

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 02:14:16 / Search time 1264.21 Seconds  
(without alignments)  
2383.909 Million cell updates/sec

Title: US-09-509-779-3\_COPY\_141\_264

Perfect score: 124

Sequence: 1 CGATACGTGCGCATCTGCA.....ATTCCTCCACACATGCTGC 124

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_estchum:\*  
3: em\_estcin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fut:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vr1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	100.0	282	12	BM694092
2	124	100.0	360	12	BM705593
3	124	100.0	406	12	BM700705
4	124	100.0	416	9	AA258876

5	124	100.0	430	14	W38711
6	124	100.0	438	9	AA448759
7	124	100.0	460	9	AL702182
8	124	100.0	463	14	CA948789
9	124	100.0	469	9	AA010517
10	124	100.0	495	10	AM879993
11	124	100.0	531	13	BQ082699
12	124	100.0	542	12	BM761983
13	124	100.0	595	10	BF033587
14	124	100.0	620	12	BI668630
15	124	100.0	620	12	BM761711
16	124	100.0	622	13	BQ186247
17	124	100.0	622	14	CB217926
18	124	100.0	629	13	BQ632594
19	124	100.0	641	13	BQ632301
20	124	100.0	651	10	BM714665
21	124	100.0	653	13	BQ665507
22	124	100.0	654	12	BM979431
23	124	100.0	678	13	BU601136
24	124	100.0	702	9	AI338342
25	124	100.0	706	12	BM704660
26	124	100.0	715	10	BM718227
27	124	100.0	716	12	BM761806
28	124	100.0	745	10	BM712625
29	124	100.0	748	10	BM576768
30	124	100.0	754	12	BM766992
31	124	100.0	767	12	BI601470
32	124	100.0	793	10	BM111145
33	124	100.0	797	12	BI832643
34	124	100.0	797	12	BI858307
35	124	100.0	808	12	BM708518
36	124	100.0	815	12	BI457840
37	124	100.0	822	12	BI668735
38	124	100.0	827	12	BI828930
39	124	100.0	840	10	BM037022
40	124	100.0	847	12	BM773490
41	124	100.0	853	13	BQ876591
42	124	100.0	855	10	BE747000
43	124	100.0	856	10	BM753323
44	124	100.0	856	12	BI601855
45	124	100.0	864	12	BI757881

## ALIGNMENTS

RESULT 1  
LOCUS BM694092 282 bp mRNA linear EST 28-FEB-2002  
DEFINITION UI-E-CII-af-o-15-0-UI-r1 UI-E-CII Homo sapiens cDNA clone  
UI-E-CII-af-o-15-0-UI 5', mRNA sequence.

ACCESSION BM694092.1 GI:19007350  
VERSION  
KEYWORDS  
SOURCE

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 282)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and Subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE  
97044477

## COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

# FEATURES

source

Location/Qualifiers  
 1. 282  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-C11-af-o-15-0-UI"  
 /tissue\_type="RPE and Choroid"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-C11"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-C11 is a normalized cDNA library containing the  
 following tissue(s): RPE and Choroid. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is ACCTA.  
 This library was created for the program, Gene Discovery  
 in the Visual System, supported by National Eye Institute  
 (NEI)."

BASE COUNT 66 a 66 c 92 g 57 t 1 others  
 ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-54;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCCATCTGCAGGGTCCAGTGATGATGCTCTTAAATGCAAGCTGA 60  
 DB 137 CGATACGTGGCCCATCTGCAGGGTCCAGTGATGATGCTCTTAAATGCAAGCTGA 196  
 QY 61 AAACAACAAGAGAGCTGTGTGGTCTGGGAGAAATGATATCTTCTCCCAACTG 120  
 DB 197 AAACAACAAGAGAGCTGTGTGGTCTGGGAGAAATGATATCTTCTCCCAACTG 256  
 QY 121 CTGC 124  
 DB 257 CTGC 260

RESULT 2 360 bp mRNA linear EST 28-FEB-2002  
 BM705593  
 LOCUS  
 DEFINITION UI-E-DW0-agi-c-06-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
 ACCESION BM705593  
 VERSION BM705593.1 GI:19018851  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 360)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics

University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

# FEATURES

source

Location/Qualifiers  
 1. 360  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agi-c-06-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT73-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGATTACCGA. This library  
 was created for the program, Gene Discovery in the Visual  
 System, supported by National Eye Institute (NEI)."

BASE COUNT 90 a 79 c 106 g 85 t  
 ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 3e-54;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCCATCTGCAGGGTCCAGTGATGATGCTCTTAAATGCAAGCTGA 60  
 DB 54 CGATACGTGGCCCATCTGCAGGGTCCAGTGATGATGCTCTTAAATGCAAGCTGA 113  
 QY 61 AAACAACAAGAGAGCTGTGTGGTCTGGGAGAAATGATATCTTCTCCCAACTG 120  
 DB 114 AAACAACAAGAGAGCTGTGTGGTCTGGGAGAAATGATATCTTCTCCCAACTG 173  
 QY 121 CTGC 124  
 DB 174 CTGC 177

RESULT 3 406 bp mRNA linear EST 28-FEB-2002  
 BM700705  
 LOCUS  
 DEFINITION UI-E-CR1-aei-c-03-0-UI.r1 UI-E-CR1 Homo sapiens cDNA clone  
 ACCESION BM700705  
 VERSION BM700705.1 GI:19013963  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 406)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

FEATURES  
source

Location/Qualifiers  
1. 406  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-CRI-acl-c-03-0-UI"  
/tissue\_type="eye anterior segment"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-CRI"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CRI is a normalized cDNA library containing the following tissue(s): eye anterior segment. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (GT)18 tail. The sequence tag for this library is AATCGCGCAT. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT  
ORIGIN

90 a 104 c 133 g 79 t

Query Match  
Best Local Similarity 100.0%; Score 124; DB 12; Length 406;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTGTCTTAGATGTCACACTGA 60  
|||||  
Db 178 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTGTCTTAGATGTCACACTGA 237

QY 61 AAACAACAAGAGAGACTGTGTGTGTGTGTGGGAGATGATGATTCCTTCCACAACTG 120  
|||||  
Db 238 AAACAACAAGAGAGACTGTGTGTGTGTGTGGGAGATGATGATTCCTTCCACAACTG 297

QY 121 CTGC 124  
|||||  
Db 298 CTGC 301

RESULT 4  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA258876 416 bp mRNA linear EST 13-AUG-1997  
AA258876  
similar to WP:ZK287.5 CB06614 ; mRNA sequence.  
AA258876.1 GI:1894001  
EST.  
Homo sapiens (human)  
Homo sapiens

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 416)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 795 Std Error: 0.00  
Seq primer: -28mJ rev2 ET from Amersham.

FEATURES  
source

Location/Qualifiers  
1. 416  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:686856"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GCBI"  
/note="Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Altman (NCI) and Dr. Gerald Marcu (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCACTGTAAGAGGAGGAGGCGCCGCTCATTTTTTTTTTTT-3'  
(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
ORIGIN

102 a 97 c 125 g 92 t

Query Match  
Best Local Similarity 100.0%; Score 124; DB 9; Length 416;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTGTCTTAGATGTCACACTGA 60  
|||||  
Db 102 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTGTCTTAGATGTCACACTGA 161

QY 61 AAACAACAAGAGAGACTGTGTGTGTGTGTGGGAGATGATGATTCCTTCCACAACTG 120  
|||||  
Db 162 AAACAACAAGAGAGACTGTGTGTGTGTGTGGGAGATGATGATTCCTTCCACAACTG 221

QY 121 CTGC 124  
|||||  
Db 222 CTGC 225

RESULT 5  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

W38711 430 bp mRNA linear EST 15-MAY-1996  
W38711  
2b27c08.r1 Soares parathyroid tumor NBHPA Homo sapiens cDNA clone  
IMAGE:304814 5' similar to WP:F35612.9 CE00978 ; mRNA sequence.  
W38711.1 GI:1320487  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 430)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maira, M., Parsons, J., Riekin, L., Rohlfing, T., Soares, M., Tan, F., Trevasaki, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck Est Project





Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by GFP (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No SI sequence available.  
This clone (DKFZp68606155) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source

1.460

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp68606155"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_1ib="686 (synonym: h1cc3)"  
/note="Vector: pTriblex2; Site\_1: SfilA; Site\_2: SfilB; cDNA-collection"

BASE COUNT 105 a 114 c 148 g 93 t  
ORIGIN

Query Match 100.0%; Score 124; DB 9; Length 460;  
Best Local Similarity 100.0%; Pred. No. 3.2e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGGCATCTGCAGGCTCAGATGATGATGCTCTTATGATGCAAGCTGA 60  
166 CGATACGTGCGGCATCTGCAGGCTCAGATGATGATGCTCTTATGATGCAAGCTGA 225  
DB 61 AAACAAACAGAGAGACTGTGTGCTGGGGAGATGATGATCATCTCTCCACAACTG 120  
226 AAACAAACAGAGAGACTGTGTGCTGGGGAGATGATGATCATCTCTCCACAACTG 285  
QY 121 CTGC 124  
111  
DB 286 CTGC 289

RESULT 8  
LOCUS CA948789 463 bp mRNA linear EST 31-DEC-2002  
DEFINITION iQ27g05.Y1 HR85 islet Homo sapiens cDNA clone IMAGE: 5' similar to  
TR:Q9Y5M7 Q9Y5M7 RING FINGER PROTEIN. [1]; mRNA sequence.  
ACCESSION CA948789  
VERSION CA948789.1 GI:27441666  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 463)  
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K., Lemishka, I., Scares, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Maria, M., Pepe, D., Wyllie, T., Martin, J., Blistein, A., Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, W., McCann, R., Cole, R., Tsagaris, V., Williams, T., Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished  
Other ESTs: iQ27g05.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biochem.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -400P from Gibco  
High quality sequence stop: 394.  
Location/Qualifiers

## FEATURES

source

1.463

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_1ib="HR85 islet"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1: NotI; Site\_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permut Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

BASE COUNT 114 a 98 c 132 g 119 t  
ORIGIN

Query Match 100.0%; Score 124; DB 14; Length 463;  
Best Local Similarity 100.0%; Pred. No. 3.2e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGGCATCTGCAGGCTCAGATGATGATGCTCTTATGATGCAAGCTGA 60  
66 CGATACGTGCGGCATCTGCAGGCTCAGATGATGATGCTCTTATGATGCAAGCTGA 125  
DB 61 AAACAAACAGAGAGACTGTGTGCTGGGGAGATGATGATCATCTCTCCACAACTG 120  
126 AAACAAACAGAGAGACTGTGTGCTGGGGAGATGATGATCATCTCTCCACAACTG 185  
QY 121 CTGC 124  
111  
DB 186 CTGC 189

RESULT 9  
LOCUS AA010517 469 bp mRNA linear EST 09-MAY-1997  
DEFINITION z109g06.r1 Soares fetal\_liver\_spleen\_infls\_S1 Homo sapiens cDNA  
clone IMAGE:430330 5' similar to WP:F35612.9 CE00978; mRNA  
sequence.  
ACCESSION AA010517  
VERSION AA010517.1 GI:1471543  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 469)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chisoso, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellberg, K., Soares, M.B., Tan, F., Thierly, M., Trevaetis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
8889549  
Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the



source

1. .531  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S14K402-56-D03"  
/cell\_line="K402"  
/lab\_host="Top10F"  
/clone\_1ib="S14K402"  
/note="Organ: Stomach; Vector: pTZ18Rpl; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A) + RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 120 a 141 c 161 g 109 t  
ORIGIN

Query Match 100.0%; Score 124; DB 13; Length 531;  
Best Local Similarity 100.0%; Pred. No. 3.4e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCAGACTGA 60  
DB 205 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCAGACTGA 264

QY 61 AAACAACAAGAGAGACTGTGTGTGCTCTGGGAGAGATGATATCATCTCTCCACAACTG 120  
DB 265 AAACAACAAGAGAGACTGTGTGTGCTCTGGGAGAGATGATATCATCTCTCCACAACTG 324

QY 121 CTGC 124  
DB 325 CTGC 328

RESULT 12  
BM761983 542 bp mRNA linear EST 04-MAR-2002  
LOCUS K-EST0043062 S13KMS5 Homo sapiens cDNA clone S13KMS5-5-A10 5', mRNA  
DEFINITION sequence.  
ACCESSION BM761983  
VERSION BM761983.1 GI:19091598  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 542)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished  
CONTACT: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boseon-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 5 row: A column: 10  
High quality sequence stop: 542.  
Location/Qualifiers  
1. .542

FEATURES  
source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S13KMS5-5-A10"  
/tissue\_type="myeloma"  
/cell\_line="KMS-5"  
/lab\_host="Top10F"  
/clone\_1ib="S13KMS5"  
/note="Vector: pCMS; Site\_1: EcoRI; Site\_2: NotI; The poly  
(A) + RNA was dephosphorylated with bacterial alkaline  
phosphatase (BAP) and then decapped with tobacco acid  
pyrophosphatase (TAP). The decapped intact mRNA was  
ligated with DNA-RNA linker including EcoR I site by  
treatment of T4 RNA ligase and the first strand cDNA was  
synthesized from oligo dt-selected mRNA by priming with  
dt-tailed vector. The dt-tailed vector was adjusted to  
have about 60nt. The cDNA vector was circularized with E.  
coli DNA ligase after digestion of EcoRI which site is  
also included in vector. An RNA strand converted to a DNA  
strand by Okayama-Berg method. The obtained cDNA vectors  
were used for transformation of competent cells E. coli  
Top10F by electroporation method. The cDNA libraries  
constructed by this method are full-length enriched cDNA  
library."

BASE COUNT 125 a 137 c 164 g 116 t  
ORIGIN

Query Match 100.0%; Score 124; DB 13; Length 542;  
Best Local Similarity 100.0%; Pred. No. 3.4e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCAGACTGA 60  
DB 182 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCAGACTGA 241

QY 61 AAACAACAAGAGAGACTGTGTGTGCTCTGGGAGAGATGATATCATCTCTCCACAACTG 120  
DB 242 AAACAACAAGAGAGACTGTGTGTGCTCTGGGAGAGATGATATCATCTCTCCACAACTG 301

QY 121 CTGC 124  
DB 302 CTGC 305

RESULT 13  
BF033587 595 bp mRNA linear EST 20-OCT-2000  
LOCUS 601453564P1 NIH\_MGC\_66 Homo sapiens cDNA clone IMAGE:3857511 5',  
DEFINITION mRNA sequence.  
ACCESSION BF033587  
VERSION BF033587.1 GI:10741299  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 595)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
CONTACT: Robert Strusberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DBP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
http://image.llnl.gov  
Plate: L1AM9587 row: 0 column: 16  
High quality sequence stop: 595.  
Location/Qualifiers  
1. .595

FEATURES  
source

	/organism="Homo sapiens"
	/mol_type="mrna"
	/db_xref="taxon:9606"
	/clone="IMAGE:3857511"
	/tissue_type="adenocarcinoma"
	/lab_host="DH10B (phage-resistant)"
	/clone_lib="NIH MGC 66"
	/note="Organ: ovary; Vector: PCMV-SPORE6; Site_1 NotI;
	Site_2 SalI; Cloned unidirectionally. Primer: Oligo dT.
	Average insert size 1.8 kb. Library constructed by Life
	Technologies."
BASE COUNT	153 a 126 c 165 g 151 t
ORIGIN	
Query Match	100.0%; Score 124; DB 10; Length 595;
Best Local Similarity	100.0%; Pred.No. 3;Se-54;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1 CGATACGTGCGCCATCTGCAGGGTCAGAGTGATGATGCTGCTGCTTAGATGTCAAGCTGA 60
Db	96 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTGCTGCTTTAGATGTCAAGCTGA 155
QY	61 AAACAACAAGAAGACTGTGTTGTGGTCTGGGGAGAGATTAATCAATCCTTCACAACTG 120
Db	156 AAAACAACAAGAAGACTGTGTTGTGGTCTGGGGAGAGATTAATCAATCCTTCACAACTG 215
QY	121 CTGC 124
Db	216 CTGC 219

RESULT	14
LOCUS	B1668630
DEFINITION	B1668630 600 bp mRNA linear EST 12-SEP-2001 603293843P1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5313034 5', mRNA sequence.
ACCESSION	B1668630
VERSION	B1668630.1 GI:15528263
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 600) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: cgapbs@email.nih.gov
TITLE	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
JOURNAL	cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (Riken)
COMMENT	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLM11793 row: f column: 11 High quality sequence stop: 596. Location/Qualifiers
FEATURES	

```

FEATURES
source
    location/Qualifiers
    1. 600
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:5313034"
        /tissue_type="hypothalamus"
        /lab_host="DH10B"
        /clone_lib="NIH_MGC_96"
        /note="Organ: brain; Vector: pBluescript (modified
        pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
        ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
        size-selected for average insert size 2.3 kb and
        normalized to RQF 5. This is a primary library enriched

```

[illegible]

FEATURES	REFERENCE	ORGANISM	DEFINITION	LOCUS	RESULT 15
BM767171	Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Kim, Y.S.	Homo sapiens (human)	620 bp mRNA linear EST 04-MAR-2002	BM767171	
BM767171	Oh, J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.	Homo sapiens	EST	BM767171	
BM767171.1	21C Frontier Korean EST Project 2001	Unpublished	High quality sequence stop: 620.	BM767171.1	
GI:19096786	Genome Research Center	Contact: Kim YS	Location/Qualifiers	GI:19096786	
EST.	Korea Research Institute of Bioscience & Biotechnology			EST	
sequence.	52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea			sequence.	
	Tel: +82-42-860-4470				
	Fax: +82-42-860-4409				
	Email: yongsung@mail.kribb.re.kr				
	Plate: 4 row: C column: 07				

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S14K402-4-C07"
/cell_line="K402"
/lab_host="Top10F"
/clone_lib="S14K402"
/notes="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of

```

competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 145 a 154 c 177 g 144 t

Query Match 100.0%; Score 124; DB 12; Length 620;  
Best Local Similarity 100.0%; Pred. No. 3.5e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTCGCCATCTGCAGGGTCCAGTGATGATGCTGCTTAGATGCAACTGA 60  
DB 205 CGATACGTCGCCATCTGCAGGGTCCAGTGATGATGCTGCTTAGATGCAACTGA 264  
QY 61 AAACAACAAGAGAGACTGTGTGTGTCGGGGAGATGATTAATCAATCTTCCACAACCTG 120  
DB 265 AAACAACAAGAGAGACTGTGTGTGTCGGGGAGATGATTAATCAATCTTCCACAACCTG 324  
QY 121 CTGC 124  
DB 325 CTGC 328

RESULT 16 BQ186247 621 bp mRNA linear EST 30-APR-2002  
LOCUS BQ186247  
DEFINITION UI-E-EJ1-aj1-1-03-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone  
ACCESSION BQ186247  
VERSION BQ186247.1 GI:20361798  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 621)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newson Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 REVERSE

FEATURES  
source Location/Qualifiers

1..621  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-EJ1-aj1-1-03-0-UI"  
/tissue\_type="fetal eyes, lens, eye anterior segment,  
optic nerve, retina, Retina Foveal and Macular, RPE and  
Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_1ib="UI-E-EJ1"  
/note="Organ: eye, Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Ecor I; Site\_2: Not I;  
UI-E-EJ1 is a subtracted cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an Ecor I adaptor, digested  
with Not I, and cloned directionally into pT73-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are: fetal eyes, AGAATCAACA  
; lens, CGATTAGCGA; eye anterior segment, AATGCCCAT;  
optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and  
Macular, GTCC; RPE and Choroid, ACCCTA. This library was  
created for the program, Gene Discovery in the Visual  
System, supported by National Eye Institute (NEI)."

BASE COUNT 168 a 121 c 160 g 170 t 2 others

Query Match 100.0%; Score 124; DB 13; Length 621;  
Best Local Similarity 100.0%; Pred. No. 3.5e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTCGCCATCTGCAGGGTCCAGTGATGATGCTGCTTAGATGCAACTGA 60  
DB 20 CGATACGTCGCCATCTGCAGGGTCCAGTGATGATGCTGCTTAGATGCAACTGA 79  
QY 61 AAACAACAAGAGAGACTGTGTGTGTCGGGGAGATGATTAATCAATCTTCCACAACCTG 120  
DB 80 AAACAACAAGAGAGACTGTGTGTGTCGGGGAGATGATTAATCAATCTTCCACAACCTG 139  
QY 121 CTGC 124  
DB 140 CTGC 143

RESULT 17 CB217926 622 bp mRNA linear EST 06-FEB-2003  
LOCUS CB217926  
DEFINITION NISC nb05g10.y1 COGENE 6E MAN Homo sapiens cDNA clone IMAGE:5795514  
ACCESSION CB217926  
VERSION CB217926.1 GI:28266118  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 622)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
cDNA Library Preparation:  
DNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
info@image.llnl.gov  
Plate: LLM12898 row: M column: 19  
Seq primer: M13P1 reverse primer (ABI).  
Location/Qualifiers

FEATURES  
source

1..622  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5795514"  
/tissue\_type="mandible, pooled"  
/dev\_stage="embryo, 6 weeks postconception"  
/lab\_host="DH10B"  
/clone\_1ib="COGENE 6E MAN"  
/note="Vector: pAMP1; cDNA primed using oligo-dT primer,  
directionally cloned into UDG sites of pAMP1. Size  
selected for insert sizes ranging from 0.2-2.0 kb.

Normalized to Coe5. Primary library, non-amplified. Library constructed by M. Lovett. For more information on this library, please contact R. Tiedell (Washington University) or visit the COGENE website at <http://www.cogene.org>

BASE COUNT	144 a	147 c	182 g	149 f
ORIGIN				

Query Match	Score	DB	Length
100.0%	124	14	622

Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGATACCTGGCCATCTGCAGGGTCCAGGTGTGATGCGCTCTTAATATCTCAAGCTGA	60
Db	165	CGATACCTGGCCATCTGCAGGGTCCAGGTGTGATGCGCTCTTAATATCTCAAGCTGA	224
QY	61	AAACAACAGAGAGACTGTTGTGGTCTGGGAGATGTATCATTCCTCCACAACCTG	120
Db	225	AAACAACAGAGAGACTGTTGTGGTCTGGGAGATGTATCATTCCTCCACAACCTG	284
QY	121	CTGC 124	
Db	285	CTGC 288	

RESULT 18  
BQ632594

LOCUS	623 bp	mRNA	linear	EST 02-J
DEFINITION	BO632594			
	i125c04.y1	HR85 1stet Homo sapiens cDNA clone IMAGE:6031086		
	similar to	TR:Q9Y5M7	Q9Y5M7	RING FINGER PROTEIN. [1] ; mRNA
	sequence.			

ACCESSION	BQ632594	GI:21684112
VERSION	BQ632594.1	
KEYWORDS	EST.	

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Cetartihii; Homnidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 629)  
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,  
Lemishka, I., Searce, M., Bresciani, J., Gradow, G., Clifton, S.,  
Hillier, L., Marx, M., Page, D., Wylie, T., Martin, J., Bilsland, A.,  
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas  
, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvilli, R., Williams, T  
, Jackson, Y., and Bowers, Y.

**FEATURES**  
**source**

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="eaxon:9606"
/clone="IMAGE:603106"
/tissue_type="Purified pancreatic islet"
/lab_host="RDH10P"
/clone_lib="HR85 islet"
/notes="Organ: Pancreas; Vector: pBluescript SK(-); Site 1
      _2: XhoI; cDNA made by oligo-dT priming."

```

Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permut lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@wustl.edu, Tel: 314-362-1916 Fax: 314-747-2692."

BASE COUNT	157 a	141 c	173 g	158 c
ORIGIN				

Query Match	100.0%;	Score 124;	DB 13;	Length 629;
Best Local Similarity	100.0%;	Pred. No. 3.5e-54;		
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	CGATTAAGTGGCCATCTCTCAGGGATC	CAGAGTCATGAGTCCCTGCTTTCAGATGCTCAACTGA	60
Db	126	CGATACGTGGCCATCTCTCAGGGATC	CAAGTCATGAGTCCCTGCTTTCAGATGCTCAACTGA	185
Qy		AAAACAAACAAGAGACTGTGTTGTGCTCGGGGAGAAATGTAATCATTCCTCCACAACATG		120
Db	61	AAAACAAACAAGAGACTGTGTTGTGCTCGGGGAGAAATGTAATCATTCCTCCACAACATG		120
Qy		AAACAAACAACAAGAGACTGTGTTGTGCTCGGGGAGAAATGTAATCATTCCTCCACAACATG		245
Db	186	AAACAAACAACAAGAGACTGTGTTGTGCTCGGGGAGAAATGTAATCATTCCTCCACAACATG		245
Qy		CTGC	124	
Db	246	CTGC	249	

RESULT 19					
BQ632301/c					
LOCUS	BQ632301	641 bp	mRNA	linear	EST 02-JUL-2002
DEFINITION	1125C04.x1 HR85 islet Homo sapiens cDNA IMAGE:6031086 3'				

similar to TR:Q9Y5M7 Q9  
sequence.  
BQ632301

VERSION	BQ632301.1	GI:21683819
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
Melton, D., Brown, J., Keny, G., Pernutt, A., Lee, C., Kaestner, K.,  
1 (bases 1 to 641)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 641)

**TITLE** Endocrine Pancreas Consortium  
**JOURNAL** Unpublished  
**COMMENT** Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

## FEATURES

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6031086"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_id="NH85 islet"
note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1

```

NotI; Site 2; XhoI; cDNA made by oligo-dT priming.  
Size-selected on agarose gel. Average insert size ~1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permut Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

BASE COUNT  
ORIGIN

176 a 159 c 125 g 181 t

Query Match 100.0%; Score 124; DB 13; Length 641;  
Best Local Similarity 100.0%; Pred. No. 3.5e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATAGTGGCCATCTGCAGGGTCCAGTGAATGCTGCTTGAATGTCACACTGA 60  
|||||  
Db 624 CGATAGTGGCCATCTGCAGGGTCCAGTGAATGCTGCTTGAATGTCACACTGA 565

QY 61 AACCAACAAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 120  
|||||  
Db 564 AACCAACAAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 505

QY 121 CTGC 124  
|||||  
Db 504 CTGC 501

RESULT 20  
LOCUS BG714665 651 bp mRNA linear EST 08-MAY-2001  
DEFINITION 602677016F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4799834 5',  
mRNA sequence.

ACCESSION BG714665  
VERSION BG714665.1 GI:13993596  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1. (bases 1 to 651)  
NIH-MGC http://mgi.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapds@mail.nih.gov  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHRG), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LHAM10689 row: 0 column: 03  
High quality sequence stop: 643.

## FEATURES

source

1. 651  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4799834"  
/issue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_96"  
/note="Organ: Brain; Vector: pBluescript (modified  
bluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.3 kb and  
normalized to 10<sup>5</sup>. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIMH/NHRG), National

Institutes of Health). Note: this is a NIH\_MGC library."

BASE COUNT 157 a 148 c 190 g 156 t  
ORIGIN

Query Match 100.0%; Score 124; DB 10; Length 651;  
Best Local Similarity 100.0%; Pred. No. 3.5e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATAGTGGCCATCTGCAGGGTCCAGTGAATGCTGCTTGAATGTCACACTGA 60  
|||||  
Db 169 CGATAGTGGCCATCTGCAGGGTCCAGTGAATGCTGCTTGAATGTCACACTGA 228

QY 61 AACCAACAAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 120  
|||||  
Db 229 AACCAACAAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 288

QY 121 CTGC 124  
|||||  
Db 289 CTGC 292

RESULT 21  
LOCUS BU665507 653 bp mRNA linear EST 30-SEP-2002  
DEFINITION c132b08.z1 Hembase; Erythroid Precursor Cells (LCB:c1 library)  
Homo sapiens cDNA clone c132b08 5', mRNA sequence.

ACCESSION BU665507  
VERSION BU665507.1 GI:23377694  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1. (bases 1 to 653)  
Gubin, A.N., Lee, Y.T., Bouffard, G.G. and Miller, J.L.  
Gene Expression in Human Erythroid Precursor Cells  
Unpublished

AUTHORS Laboratory of Chemical Biology  
TITLE Laboratory of Chemical Biology  
JOURNAL  
COMMENT Contact: Jeffery L. Miller  
Laboratory of Chemical Biology  
National Institute of Diabetes and Digestive and Kidney Diseases  
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD  
20892, USA  
Tel: 301 402 2373  
Fax: 301 435 5148  
Email: jml7@nih.gov

The 'c1' library was constructed by Alexander Gubin, Ph.D. in the  
Laboratory of Chemical Biology, NIDDK, NIH. DNA Sequencing and/or  
analyses by National Institutes of Health Intramural Sequencing  
Center (NISC). More information available at:  
http://hembase.nidk.nih.gov  
Plate: 132 row: 0 column: 08  
Seq primer: 5' lambda-TripLex2 Sequencing Primer.  
Location/Qualifiers

## FEATURES

source

1. 653  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="c132b08"  
/sex="unknown"  
/issue\_type="Blood"  
/cell\_line="Erythroid Precursor Cells"  
/cell\_line="Primary Culture of Peripheral Blood  
Mononuclear Cells"  
/dev\_stage="Precursor erythroblasts; GPA++"  
/lab\_host="DH5alpha"  
/clone\_lib="Hembase; Erythroid Precursor Cells (LCB:c1  
library)"  
/note="Organ: blood; Vector: pTriplex; Site 1: SfiI;  
Site 2: SfiI; A complementary DNA (cDNA) library from  
human erythroid precursor cells was constructed using  
SMART PCR (polymerase chain reaction) cDNA library  
construction kit (Clontech, Palo Alto, CA) according to  
the manufacturer's directions, but with slight

modifications. Briefly, reverse transcription was performed in the presence of 1 umol/l peptide nucleic acid (PNA) oligos (N-terminal )-biotin-GTC-CAC-CCG-AGG-CCT-G-(C-terminal) and (N-terminal) -biotin-C(T/C)T-GAA-GTT-CTC-AGG-A-(C-terminal) . Synthesized cDNA was digested with SfiI and size-selected on a 1% agarose gel (>800bp). Large-scale sequencing of the library was performed by the NIH Intramural Sequencing Center (NISC; <http://www.nisc.nih.gov/>).

BASE COUNT 147 a 163 c 187 g 156 t  
ORIGIN

Query Match 100.0%; Score 124; DB 13; Length 653;  
Best Local Similarity 100.0%; Pred. No. 3.6e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCATCTGCAGGCTCCAGTGGATGCTGCTTAGATGTAAGCTGA 60  
DB 220 CGATACGTGGCCATCTGCAGGCTCCAGTGGATGCTGCTTAGATGTAAGCTGA 279  
QY 61 AAACAAACAGAGAGACTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 120  
DB 280 AAACAAACAGAGAGACTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 339  
QY 121 CTGC 124  
DB 340 CTGC 343

RESULT 22  
BM979431 654 bp mRNA linear EST 21-FEB-2003  
LOCUS  
DEFINITION UI-CF-DUI-adr-f-23-0-UI-s1 UI-CF-DUI Homo sapiens cDNA clone  
BM979431  
UI-CF-DUI-adr-f-23-0-UI 3', mRNA sequence.  
BM979431  
EST.  
GI:19599873

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 654)  
Bonald,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL  
MEDLINE  
PUBMED  
Genome Res. 6 (9), 791-806 (1996)  
97044477

COMMENT  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics ([www.researchgen.com](http://www.researchgen.com)) or from Open Biosystems  
([www.openbiosystems.com](http://www.openbiosystems.com)).  
Seq primer: M13 FORWARD  
POLYA=Yes

FEATURES  
SOURCE  
1. 654  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI-adr-f-23-0-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"

/lab host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone.lib="UI-CF-DUI"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-CF-DUI is a normalized cDNA library containing the  
following tissue(s): Primary Lung Epithelial Cells The  
library was constructed according to Bonald, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT73-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (drr)18 tail. The sequence tag for this  
library is GGCTGAGGC.  
TAG\_LIB=UI-CF-DUI  
TAG\_TISSUE=Lung Epithelial Cells Tissue nos 359-368  
TAG\_SEQ=GGCTGAGGC"

BASE COUNT 173 a 156 c 127 g 198 t

Query Match 100.0%; Score 124; DB 12; Length 654;  
Best Local Similarity 100.0%; Pred. No. 3.6e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCATCTGCAGGCTCCAGTGGATGCTGCTTAGATGTAAGCTGA 60  
DB 643 CGATACGTGGCCATCTGCAGGCTCCAGTGGATGCTGCTTAGATGTAAGCTGA 584  
QY 61 AAACAAACAGAGAGACTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 120  
DB 583 AAACAAACAGAGAGACTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 524  
QY 121 CTGC 124  
DB 523 CTGC 520

RESULT 23  
BU601136 678 bp mRNA linear EST 20-SEP-2002  
LOCUS  
DEFINITION AGENCOURT 10018944 NIH MGC\_142 Homo sapiens cDNA clone  
IMAGE:6495009 5', mRNA sequence.  
BU601136  
IMAGE:6495009 5', mRNA sequence.  
BU601136  
EST.  
GI:23252895

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 678)  
NIH-MGC <http://imgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
MEDLINE  
PUBMED  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: NCI  
CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM2672 row: c column: 10  
High quality sequence drop: 527.

FEATURES  
SOURCE  
1. 678  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6495009"  
/tissue\_type="mixed (pool of 40 RNAs)"



/lab host="DH10B-T1-phage-resistant")  
/clone \_lib="NIH\_MGC\_142"  
/notes="Vector: pDNR-LIB; Site\_1: Sfil (ggccatttgcc); Site\_2: Sfil (ggccctgcgc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:  
5'-ATCTAGAGCGCCGAGCGCGCCGACATC-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART Kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH MGC 141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."  
BASE COUNT 185 a 129 c 160 g 178 t 26 others  
ORIGIN  
Query Match 100.0%; Score 124; DB 13; Length 678;  
Best Local Similarity 100.0%; Pred. No. 3.6e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATACGTCGCCCATTCGACGGGTCACAGTATGATGATCCGCTTATGATGTCACGCTGA 60  
|||||  
13 CGATACGTCGCCCATTCGACGGGTCACAGTATGATGATCCGCTTATGATGTCACGCTGA 72  
QY 61 AAACAAACAGAGGACGTGTGTGGGTCTGGGGGAAGATGATATATTCCTCCCAACTG 120  
|||||  
73 AAACAAACAGAGGACGTGTGTGGGTCTGGGGGAAGATGATATATTCCTCCCAACTG 132  
QY 121 CTGC 124  
|||||  
DB 133 CTGC 136  
RESULT 24  
LOCUS A1338342/c  
DEFINITION 702 bp mRNA linear EST 13-FEB-1999  
q966903.x1 Soares total fetus Nb2HP8.9w Homo sapiens cDNA clone IMAGE:193252 3' similar to Wf:R10A10.2 CE12670 ZINC FINGER PROTEIN ;, mRNA sequence.  
ACCESSION A1338342  
VERSION A1338342.1 GI:4075269  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 702)  
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 854 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 455.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:193252"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares total\_fetus\_Nb2HP8.9w"  
/notes="Vector: pT7TD-Pac (Pharmacia) with a modified

	polylinker; Site_1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGGAGCGCCGTAAATTATTTTTTTTTT 3'].
	Totally-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	189 a 175 c 137 g 201 t
ORIGIN	
Query Match	100.0%; Score 124; DB 9; Length 702;
Best Local Similarity	100.0%; Pred. No. 3,6e-54;
Matches 124; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	CGATACGGCGGCCCATCTGCAGGGTCAGGGATGATGATCCTTGATGTAAGCTGA 60       CGATACGTGCCCATCTGCAGGGTCAGGGATGATGATGATCCTTGATGTAAGCTGA 569
Oy	61 AAACAACAAGAGACTGTGTGTTGTGTCTGGGAGAATGTATCATCTCTTCCAACACTG 120 
Dd	568 AAAACAACAAGAGACTGTGTGTTGTGTCTGGGAGAATGTATCATCTCTTCCAACACTG 509
Oy	121 CTGC 124 
Db	508 CTGC 505
RESULT 25	
BMT04660	
LOCUS	BMT04660 706 bp mRNA linear EST 28-FEB-2002
DEFINITION	UI-E-CII-afo-o-15-O-UI.r2 UI-E-CII Homo sapiens CDNA clone
ACCESSION	BMT04660
VERSION	BMT04660.1 GI:19017918
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 706)	
Bonaldo,M.F., Lennon,G. and Soares,M.B.	
Normalization and subtraction: two approaches to facilitate gene discovery	
Genome Res. 6 (9), 791-806 (1996)	
JOURNAL MEDLINE PUBMED COMMENT	Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA Tel.: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu Tissue Procurement: Dr. Gregg Hageman CDNA library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Reverse. Location/Qualifiers 1..706 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="UI-E-CII-afo-o-15-0-UI" /tissue type="RPE and Choroid" /dev stage="adult" /lab_host="DHIOB (Life Technologies) (TI phage resistant)" /clone_idb="UI-E-CII" /note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a
FEATURES	
source	





ACCESSION	BG766992				
VERSION	BG766992.1	GI:14077645			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	NIH-MGC http://mgi.mcg.nih.gov/				
AUTHORS	1 (bases 1 to 754)				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC/DCTD/BTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Place: LHCN1743 row: e column: 04 High quality sequence stop: 750. Location/Qualifiers				
FEATURES	1..754				
source	/organism="Homo sapiens" /mol_type="RNA" /db_xref="taxon:9606" /clone="IMAGE:4870251" /tissue.type="melanotic melanoma, high MDR (cell line)" /lab_host="DH10B (phage-resistant)" /clone.lib="NIH MGC 49" /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC library."				
BASE COUNT	205 a	160 c	197 g	192 t	
ORIGIN					
Query Match	100.0%; Score 124;	DB 12;	Length 754;		
Best Local Similarity	100.0%; Pred. No. 3,7e-54;				
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	1 CGATACGGCGCACCTTCAGGGGTCAAGTAGTGATCGCTTTAATGTCGAAGCTGA	60			
Db	129 CGATTGCGGCCCATCTCAGGGGTCAAGGTATGATGATCCCTGTTTAATGTCAAAGCTGA	188			
OY	61 AAACAACAAGAAGAGACTGTGTTGTGTGGGGAGAAATGTAATCATTTCTTCACAACACTG	120			
Db	189 AAAACAAACAAGAAGACTGTGTTGTGTGTGGGGAGAAATGTAATCATTTCTTCACAACACTG	248			
OY	121 CTGC 124				
Db	249 CTGC 252				
RESULT 31					
LOCUS	BI601470	767 bp	mRNA	linear	EST 07-SEP-2001
DEFINITION	603249053r1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5300845 5', mRNA sequence.				
ACCESSION	BI601470				
VERSION	BI601470.1	GI:15494409			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 767)				

AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .									
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished.									
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a> Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LHAM11761 row: j column: 14 High quality sequence stop: 701. Location/Qualifiers									
FEATURES	1..767									
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5300845" /tissue_type="hypothalamus" /lab_host="DH10B" /clone_lib="NIH_MGC_96" /note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag ) ; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to R0F 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."									
BASE COUNT	192 a 173 c 214 g 188 t									
ORIGIN										
Query Match	100.0%; Score 124; DB 12; Length 767;									
Best Local Similarity	100.0%; Prid. No. 3,7e-54;									
Matches 124; Conservative	0; Mismatches 0; Indels 0; Gaps 0;									
QY	1 CGATACGCGGCCCATCTGCAGGGGTCAGGATGGATGCGCTTGATGATGTCAGGCTGA 60 									
Db	175 CGATACGCGGCCCATCTGCAGGGGTCAGGATGGATGCGCTTGATGATGTCAGGCTGA 234 									
QY	61 AAACAACAAGAGAGACTGTGTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 120 									
Db	235 AAACAACAAGAGAGACTGTGTGTGTGTGTGTGGGAGATGTAATCATTCCTCCACAAC 294 									
QY	121 CTGC 124 									
Db	295 CTGC 298 									
RESULT 32										
LOCUS	BG111145 793 bp mRNA linear EST 30-JAN-2001									
DEFINITION	60228155F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4369332 5', mRNA sequence.									
ACCESSION	BG111145									
VERSION	BG111145.1 GI:12604651									
KEYWORDS	EST.									
SOURCE	Homo sapiens (human)									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo. 1 (bases 1 to 793) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc.									

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LLM10024 row: m column: 13

High quality sequence stop: 602.

Location/Qualifiers

## FEATURES

source

1..793

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4369332"

/issue\_type="osteosarcoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_86"

/note="Organ: bone; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 200 a 187 c 218 g 187 t 1 others

ORIGIN

Query Match 100.0%; Score 124; DB 10; Length 793;

Best Local Similarity 100.0%; Pred. No. 3.7e-54;

Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCATCTGCAGGGTCCAGTGATGATGCTGCTTGAATGTCACACTGA 60

DB 157 CGATACGTGGCCATCTGCAGGGTCCAGTGATGATGCTGCTTGAATGTCACACTGA 216

QY 61 AAACAACAAGAGAGACTGTGTGTGTGGTGGGAGATGAATCATCTCTCCACAAC 120

DB 217 AAACAACAAGAGAGACTGTGTGTGTGGTGGGAGATGAATCATCTCTCCACAAC 276

QY 121 CTGC 124

DB 277 CTGC 280

RESULT 33

BI832643 797 bp mRNA linear EST 04-OCT-2001

LOCUS 603082021F1 NIH\_MGC\_120 Homo sapiens CDNA clone IMAGE:5221325 5'

DEFINITION mRNA sequence.

ACCESSION BI832643

VERSION BI832643.1 GI:15944193

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 797)

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LLM1556 row: i column: 06

High quality sequence stop: 795.

Location/Qualifiers

## FEATURES

source

1..797

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5221325"

/lab\_host="DH10B"

/clone\_1lb="NIH\_MGC\_120"

/note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH\_MGC Library."

BASE COUNT 208 a 179 c 212 g 198 t

ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 797;

Best Local Similarity 100.0%; Pred. No. 3.8e-54;

Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCATCTGCAGGGTCCAGTGATGATGCTGCTTGAATGTCACACTGA 60

DB 171 CGATACGTGGCCATCTGCAGGGTCCAGTGATGATGCTGCTTGAATGTCACACTGA 230

QY 61 AAACAACAAGAGAGACTGTGTGTGTGGTGGGAGATGAATCATCTCTCCACAAC 120

DB 231 AAACAACAAGAGAGACTGTGTGTGTGGTGGGAGATGAATCATCTCTCCACAAC 290

QY 121 CTGC 124

DB 291 CTGC 294

RESULT 34

BI858307 797 bp mRNA linear EST 10-OCT-2001

LOCUS 603384080F1 NIH\_MGC\_87 Homo sapiens CDNA clone IMAGE:5393131 5'

DEFINITION mRNA sequence.

ACCESSION BI858307

VERSION BI858307.1 GI:15999054

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 797)

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LLM12001 row: o column: 20

High quality sequence stop: 718.

Location/Qualifiers

## FEATURES

source

1..797

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5393131"

/issue\_type="mammary adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_87"

/note="Organ: breast; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT	229 a	166 c	204 g	198 t
ORIGIN				
Query Match	100.0%;	Score 124;	DB 12;	Length 797;
Best Local Similarity	100.0%;	Pred. No. 3.8e-54;		
Matches 124; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0
QY	1	CGATACGTGCGCCATCTGCAGGGGTCAGGTGATGCAATGCTGTCTGTATGATGTCGAAGCTGA	60	
Db	138	CGATACGTGCGCCATCTGCAGGGTCAGGTGATGCAATGCTGTCTGTATGATGTCGAAGCTGA	197	
QY	61	AAACAAACAAGAGAGCTGTGTGTGTGTCTGCGGGAGAAATATATCATTCCTTCCACACTG	120	
Db	198	AAACAAACAAGAGAGCTGTGTGTGTGTCTGCGGGAGAAATATATCATTCCTTCCACACTG	257	
QY	121	CTGC 124		
Db	258	CTGC 261		

RESULT 35	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
BG708518	BG708518	602670411n1 NIH_MGC_96 Homo sapiens CDNA clone IMAGE:4793091 5', mRNA sequence.	BG708518	BG708518	EST.	Homo sapiens (human)	Homo sapiens	1 (bases 1 to 805)	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLML10672 row: f column: 04 High quality sequence stop: 740.

```

FEATURES
    source
        Location/Qualifiers
            1..805
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /db_xref="IMAGE:4793091"
                /clone="IMAGE:4793091"
                /tissue_type="hypothalamus"
                /lab_host="DH10B"
                /clone_id="N1H MGC 96"
                /note="Organ: brain; Vector: pBluescript (modified
                pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgac
                ); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTTVN-3',
                size-selected for average insert size 2.3 kb and
                normalized to ROT 5. This is a primary library enriched
                for full-length clones and constructed using the
                Cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIMH/NHGRI, National
                Institutes of Health). Note: this is a N1H MGC library."
BASE COUNT
    214 a 183 c 214 g 194 t
ORIGIN
    100.0%; Score 124; DB 10; Length 805;
Query Match
    100.0%; Pred. No. 3.8e-54;
Best Local Similarity
    0; Mismatches 0; Indels 0; Gaps 0
Matches 124; Conservative

```

Qy	1	CGATACGTCGCGCATCTGAGAGGCTCCAGATGGAATGCCGTCTTGATGATCCAAAGCTGA	60
Db	166	CGATACGTCGCGCATCTGAGAGGCTCCAGATGGAATGCCGTCTTGATGATCCAAAGCTGA	225
Qy	61	AAACAACAAGAGACTGTGTGTGCTGTGGGGAAGATGATCATTCCTCCCAACTG	120
Db	226	AAACAACAAGAGACTGTGTGTGCTGTGGGGAAGATGATCATTCCTCCCAACTG	285
Qy	121	CTGC	124
Db	286	CTGC	289

RESULT 36  
BI457840

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BI457840 818 bp mRNA linear EST 21-Aug-2000  
603198212F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5277723 5',  
mRNA sequence.  
BI457840  
BI457840.1 GI:15248496  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 818)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLW)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLW at:  
<http://image.llnl.gov>  
plate: LLAM11701 row: 9 column: 04  
high quality sequence stop: 780.

	FEATURES	location/Qualifiers
	SOURCE	1. .818
OY	CGATACGTGCGCCATCTGCAGGGTCCAGTGATGANTCCTGTCTTAGATGTCGAAGCTGA	/organism="Homo sapiens"
	Best Local Similarity	/mol_type="mRNA"
	Matches 124; Conservative	/db_xref="taxon:9606"
		/clone="IMAGE:52777223"
		/issue_type="hypothalamus"
		/lab_host="DH10B"
		/clone_11b="NIH_MGC_96"
		/note="Organ: Brain; Vector: Bluescripter (modified bluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcga
		) ; Oligo-dt primed using primer 5'-TTTTTTTNNNNN-3' , and size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation) . Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT	225 a	174 c
ORIGIN	174 c	214 g
		205 t
Query Match	100.0%; Score 124;	DB 12; Length 818;
Best Local Similarity	100.0%; Pred. No. 3.8e-54;	
Matches 124; Conservative	0; Mismatches	0; Indels
		Gaps 0
OY	1 CGATACGTGCGCCATCTGCAGGGTCCAGTGATGANTCCTGTCTTAGATGTCGAAGCTGA	60
Db	157 CGATACGTGCGCCATCTGCAGGGTCCAGTGATGANTCCTGTCTTAGATGTCGAAGCTGA	216
OY	61 AAAACAACAAGAAGACTGTTGTTGCTGTGGGAGAAATGTAATCATTTCTCCCAACTG	120
Db	217 AAAACAACAAGAAGACTGTTGTTGAGTCTGGGGAGAAATGTAATCATTTCTCCCAACTG	276

QY 121 CTGC 124  
| | |  
Db 277 CTCG 280

RESULT 37  
BI668735  
LOCUS 603293174PL NIH\_MGC\_96 Homo sapiens CDNA clone IMAGE:531226 5',  
DEFINITION mRNA sequence.  
ACCESSION BI668735  
VERSION BI668735.1 GI:15582968  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 822)  
NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
AUTHORS Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11791 row: d column: 19  
High quality sequence stop: 788.  
Location/Qualifiers

FEATURES  
source  
1..822  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:531226"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_96"  
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (GTCGAG); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTT-3', size selected for average insert size 2.3 kb and normalized for ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 219 a 182 c 219 g 202 t  
ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 822;  
Best Local Similarity 100.0%; Pred. No. 3.8e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTCTTAGATGCAAGCTGA 60  
| | | | |  
Db 169 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTCTTAGATGCAAGCTGA 228  
| | | | |

QY 61 AAACAAACAAAGAGAGCTGTGTGTGTCTGGGAGAAATGTAATCTCTTCCAACTG 120  
| | | | |  
Db 229 AAACAAACAAAGAGAGCTGTGTGTGTCTGGGAGAAATGTAATCTCTTCCAACTG 288  
| | | | |

QY 121 CTGC 124  
| | |  
Db 289 CTCG 292

RESULT 38

BI828930  
LOCUS 827 bp mRNA linear EST 04-OCT-2001  
DEFINITION 603075092PL NIH\_MGC\_119 Homo sapiens CDNA clone IMAGE:5166790 5',  
mRNA sequence.  
ACCESSION BI828930  
VERSION BI828930.1 GI:15940480  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 827)  
NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
AUTHORS Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11414 row: h column: 23  
High quality sequence start: 2  
High quality sequence stop: 679.  
Location/Qualifiers

FEATURES  
source  
1..827  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5166790"  
/tissue\_type="medulla"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_119"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb. Insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH\_MGC Library."

BASE COUNT 232 a 183 c 227 g 184 t 1 others  
ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 827;  
Best Local Similarity 100.0%; Pred. No. 3.8e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTCTTAGATGCAAGCTGA 60  
| | | | |  
Db 153 CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCTCTTAGATGCAAGCTGA 212  
| | | | |

QY 61 AAACAAACAAAGAGAGCTGTGTGTGTCTGGGAGAAATGTAATCTCTTCCAACTG 120  
| | | | |  
Db 213 AAACAAACAAAGAGAGCTGTGTGTGTCTGGGAGAAATGTAATCTCTTCCAACTG 272  
| | | | |

QY 121 CTGC 124  
| | |  
Db 273 CTCG 276

RESULT 39  
BG037022 840 bp mRNA linear EST 24-JAN-2001  
DEFINITION 602287341PL NIH\_MGC\_96 Homo sapiens CDNA clone IMAGE:4374443 5',  
mRNA sequence.  
ACCESSION BG037022  
VERSION BG037022.1 GI:12432833  
KEYWORDS EST.





found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1M13595 row: h column: 04  
High quality sequence stop: 608.  
Location/Qualifiers

FEATURES  
source

1. 853  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6192819"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/clone\_lib="Lupski sympathetic trunk"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dt priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCTCAGTCTAGTCCGAGCGCCGCTT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life  
Technologies."  
BASE COUNT 231 a 191 c 225 g 205 t 1 others  
ORIGIN

Query Match 100.0%; Score 124; DB 13; Length 853;  
Best Local Similarity 100.0%; Pred. No. 3.8e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGATACGTGCGCCCATCTGCAGGGTCCAGGTGATGCTCTTAGATGCAAGCTGA 60  
DB 191 CGATACGTGCGCCCATCTGCAGGGTCCAGGTGATGCTCTTAGATGCAAGCTGA 250  
OY 61 AAACAAACAGAGAGAGCTGTGTGTGTGTGTGGAGATATATCATTCCTTCCACACTG 120  
DB 251 AAACAAACAGAGAGAGCTGTGTGTGTGTGTGGAGATATATCATTCCTTCCACACTG 310  
OY 121 CTGC 124  
DB 311 CTGC 314

RESULT 42  
BE747000 855 bp mRNA linear EST 15-SEP-2000  
LOCUS 601580743F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3929459 5',  
DEFINITION mRNA sequence.  
ACCESSION BE747000.1 GI:10160992  
VERSION BE747000  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homimidae; Homo.  
REFERENCE 1 (bases 1 to 855)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1M1762 row: e column: 12  
High quality sequence stop: 767.  
Location/Qualifiers

FEATURES  
source

1. 855  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3929459"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dt priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 250 a 161 c 238 g 205 t 1 others  
ORIGIN

Query Match 100.0%; Score 124; DB 10; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGATACGTGCGCCCATCTGCAGGGTCCAGGTGATGCTCTTAGATGCAAGCTGA 60  
DB 112 CGATACGTGCGCCCATCTGCAGGGTCCAGGTGATGCTCTTAGATGCAAGCTGA 171  
OY 61 AAACAAACAGAGAGAGCTGTGTGTGTGTGTGGAGATATATCATTCCTTCCACACTG 120  
DB 172 AAACAAACAGAGAGAGCTGTGTGTGTGTGTGGAGATATATCATTCCTTCCACACTG 231  
OY 121 CTGC 124  
DB 232 CTGC 235

RESULT 43  
BG753323 856 bp mRNA linear EST 15-MAY-2001  
LOCUS 602731740F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4875489 5',  
DEFINITION mRNA sequence.  
ACCESSION BG753323  
VERSION BG753323.1 GI:14063976  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homimidae; Homo.  
REFERENCE 1 (bases 1 to 856)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1M1756 row: o column: 10  
High quality sequence stop: 767.  
Location/Qualifiers

1. 856  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4875489"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_43"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dt priming. Directionally



OY		61	AAACAAACAAGGACTGTGGTCTGCAGGGAGAATGATCATTCCTCCACAAC	120
DB		276	AAACAAACAAGGACTGTGGTCTGCAGGGAGAATGATCATTCCTCCACAAC	335
OY		121	CTGC 124	
DB		336	CTGC 339	
RESULT 46				
CB991438				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
SOURCE				
BASE COUNT				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 124; Conservative				
OY		1	CGATACGTCGCCCACTGCAGGGTCCAGGTATGATGCTGCTTTAAGATGCAAGCTTA	60
DB		231	CGATACGTCGCCCACTGCAGGGTCCAGGTATGATGCTGCTTTAAGATGCAAGCTTA	290
OY		61	AAACAAACAAGGACTGTGGTCTGCAGGGAGAATGATCATTCCTCCACAAC	120
DB		231	AAACAAACAAGGACTGTGGTCTGCAGGGAGAATGATCATTCCTCCACAAC	350
OY		121	CTGC 124	

DB	351	CTGC	354		
RESULT 47	BG742338	904 bp	mRNA	linear	EST 15-MAY-2001
LOCUS	602631707P1	NCL_CGAP_Skn3	Homo sapiens	cDNA clone	IMAGE:4777128 5' ,
DEFINITION	mRNA sequence.				
ACCESSION	BG742338	1	GI:14052991		
VERSION	EST.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 904)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: James Clavier, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LLM10630 row: n column: 02 High quality sequence stop: 836.				
FEATURES	Location/Qualifiers				
source	1..904				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4777129"				
	/lab_host="DH10B (T1 phage-resistant)"				
	/clone_id="NCL_CGAP_Skn3"				
	/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NciI; Site 2: SalI; Cloned unidirectionally. Primer: oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCL_CGAP Library."				
BASE COUNT	215 a	223 c	254 g	211 t	1 others
ORIGIN					
Query Match	100.0%	Score 124;	DB 10;	Length 904;	
Best Local Similarity	100.0%	Pred. No. 3.9e-54;			
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	CGATACGTGCGCCATCTGCAGAGGTCAGGTCAGTGCCTGTCTTAGATGTCAGCTGA	60		
DB	274	CGATACGTGCGCCATCTGCAGAGGTCAGGTCAGTGCCTGTCTTAGATGTCAGCTGA	333		
QY	61	AAACAAACAGAGGACTGTGTTGTGTGTCGGGGAATGTATCATCTTCCACAACCTG	120		
DB	334	AAACAAACAGAGGACTGTGTTGTGTGTCGGGGAATGTATCATCTTCCACAACCTG	393		
QY	121	CTGC	124		
DB	394	CTGC	397		
RESULT 48	B1759082	918 bp	mRNA	linear	EST 25-SEP-2001
LOCUS	603042879P1	NIH_MGC_116	Homo sapiens	cDNA clone	IMAGE:5183419 5' ,
DEFINITION	mRNA sequence.				
ACCESSION	B1759082	1	GI:15750660		
VERSION	EST.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES  
source Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM11457 row: m column: 20  
High quality sequence stop: 732.  
Location/Qualifiers  
1..918  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5183419"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_116"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH MGC library."

BASE COUNT 269 a 196 c 247 g 206 t  
ORIGIN

Query Match 100.0%; Score 124; DB 12; Length 918;  
Best Local Similarity 100.0%; Pred. No. 3.9e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Gy 1 CGATACGCGCCATCTGCGAGGTCAGATGATGATCCCTCTTAATGTCGAAGCTGA 60  
Db 152 CGATACGCGCCATCTGCGAGGTCAGATGATGATCCCTCTTAATGTCGAAGCTGA 211  
Gy 61 AAACAACAAGAGACTGTGTGCTGGGAGATGATCATCTTCCTCCCAACTG 120  
Db 212 AAACAACAAGAGACTGTGTGCTGGGAGATGATCATCTTCCTCCCAACTG 271  
Gy 121 CTGC 124  
Db 272 CTGC 275

RESULT 49  
BUI92094 922 bp mRNA linear EST 04-SEP-2002  
LOCUS AGENCOURT 7968580 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6011203  
DEFINITION 5', mRNA sequence.  
ACCESSION BUI92094  
VERSION BUI92094.1 GI:22706078  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES  
source Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM13200 row: h column: 20  
High quality sequence stop: 570.  
Location/Qualifiers  
1..922  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6011203"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_68"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 249 a 202 c 250 g 220 t 1 others  
ORIGIN

Query Match 100.0%; Score 124; DB 13; Length 922;  
Best Local Similarity 100.0%; Pred. No. 3.9e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Gy 1 CGATACGCGCCATCTGCGAGGTCAGATGATGATCCCTCTTAATGTCGAAGCTGA 60  
Db 155 CGATACGCGCCATCTGCGAGGTCAGATGATGATCCCTCTTAATGTCGAAGCTGA 214  
Gy 61 AAACAACAAGAGACTGTGTGCTGGGAGATGATCATCTTCCTCCCAACTG 120  
Db 215 AAACAACAAGAGACTGTGTGCTGGGAGATGATCATCTTCCTCCCAACTG 274  
Gy 121 CTGC 124  
Db 275 CTGC 278

RESULT 50  
BG111792 947 bp mRNA linear EST 30-JAN-2001  
LOCUS 602285379F1 NIH\_MGC\_86 Homo sapiens cDNA clone IMAGE:4372819 5',  
DEFINITION mRNA sequence.  
ACCESSION BG111792  
VERSION BG111792.1 GI:12605298  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES  
source Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM10033 row: n column: 20  
High quality sequence stop: 682.  
Location/Qualifiers  
1..947  
/organism="Homo sapiens"

```
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="IMAGE:4372819"  
/tissue_type="osteosarcoma, cell line"  
/lab_host="DH10B (phage-resistant)"  
/clone_lib="NIH_MGC_86"  
/note="Organ: bone; Vector: PCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.53 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH_MGC Library." 1 others
```

BASE COUNT 228 a 241 c 267 g 210 t 1 others  
ORIGIN

Query Match 100.0%; Score 124; DB 10; Length 947;  
Best Local Similarity 100.0%; Pred. No. 3.9e-54;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCGCTGTAGATGTCAGCTGA	60
Db	194	CGATACGTGCGCCATCTGCAGGGTCCAGGTGATGATGCGCTGTAGATGTCAGCTGA	253
Qy	61	AAACAAACAGAGACTGTGTTGTTGCTGCGGAGATGTATCATTCCTCCACACTG	120
Db	254	AAACAAACAGAGACTGTGTTGTTGCTGCGGAGATGTATCATTCCTCCACACTG	313
Qy	121	CTGC	124
Db	314	CTGC	317

Search completed: November 7, 2003, 11:54:01  
Job time : 1272.21 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 11:59:37 ; Search time 2874.08 Seconds  
(without alignments)  
9640.351 Million cell updates/sec

Title: US-09-509-779-1

Perfect score: 1140  
Sequence: 1 GTTCGCGCGCGCGCATG.....ACTAATTCATCAATTAATG 1140

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1069.4	93.8	1132	11	AK003963 Mus muscu
2	1041	91.3	1114	11	AK011328 Mus muscu
3	1032.4	90.6	1084	11	AK003248 Mus muscu
4	1004.4	88.1	1078	11	AK002628 Mus muscu

5	875	76.8	997	14	BY704214	BY704214
6	856.2	75.1	948	14	BY703542	BY703542
7	836.2	73.4	989	13	BH840985	BH840985
8	833.6	73.1	901	14	CB204521	CB204521
9	802.4	70.4	1185	13	BQ964159	BQ964159
10	781.8	68.6	968	13	BQ964115	BQ964115
11	765	67.1	820	12	BT732003	BT732003
12	753.4	66.1	1079	12	BT140936	BT140936
13	740	64.9	882	14	CA980368	CA980368
14	739.6	64.9	938	14	CB209388	CB209388
15	732.2	64.2	965	14	BY710626	BY710626
16	725.4	63.6	743	14	CA323814	CA323814
17	719.4	63.1	970	13	BUS13948	BUS13948
18	707.2	62.0	728	13	BO571112	BO571112
19	700.8	61.5	722	11	AK007588	AK007588
20	687.2	60.3	721	14	BY707632	BY707632
21	686.6	60.2	731	12	BI647378	BI647378
22	682.8	59.9	889	14	CB196142	CB196142
23	680.2	59.7	867	14	CA464071	CA464071
24	673.4	59.1	978	12	BT156184	BT156184
25	671	58.9	787	14	CB318729	CB318729
26	665.4	58.4	916	12	BS917948	BS917948
27	664.6	58.3	944	14	CA977674	CA977674
28	653.4	57.3	676	13	BQ177475	BQ177475
29	651.6	57.2	686	12	BO044727	BO044727
30	651.4	57.1	754	12	BT147544	BT147544
31	651	57.1	704	10	BF018804	BF018804
32	649.2	56.9	946	10	BF83735	BF83735
33	647.8	56.8	767	14	CB571468	CB571468
34	646.2	56.7	737	12	BI556918	BI556918
35	629	55.2	642	14	BY737338	BY737338
36	627.2	55.0	920	14	CB209237	CB209237
37	624.4	54.8	822	10	BF302267	BF302267
38	612.2	53.7	755	10	BF579825	BF579825
39	597.2	52.4	620	10	BE628111	BE628111
40	596	52.3	651	9	AM555800	AM555800
41	592.4	52.0	728	14	CB951114	CB951114
42	586.8	51.5	653	14	BY757681	BY757681
43	584	51.2	624	12	BM946344	BM946344
44	582.8	51.1	757	14	CB318426	CB318426
45	580.4	50.9	779	14	CD241479	CD241479

## ALIGNMENTS

RESULT 1  
AK003963  
LOCUS 1132 bp mRNA linear HTC 05-DEC-2002  
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:110029M05 product:ring finger protein 7, full insert sequence.  
ACCESSION AK003963  
VERSION AK003963.1 GI:12834939  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
1 Carinici, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
AUTHORS Carinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20493374  
PUBMED 11042159

REFERENCE	AUTHORS	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE			RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL			Genome Res. 10 (11), 1757-1771 (2000)
PUBMED			20530913
REFERENCE	AUTHORS	4	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Alzawa, T., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabili, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, T., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schorbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.
TITLE			Functional annotation of a full-length mouse cDNA collection
JOURNAL			Nature 409 (6821), 685-690 (2001)
PUBMED			21085660
REFERENCE	AUTHORS	5	11217851
TITLE			The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team.
JOURNAL			Analysis of the mouse transcriptome based on functional annotation
PUBMED			of 60,770 full-length cDNAs
REFERENCE	AUTHORS		Nature 420, 563-573 (2002)
			5 (bases 1 to 1132)
			Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hara, A., Hara, A., Hayatsu, N., Hiramoto, K., Hirao, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komno, H., Kondo, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE			Direct Submission
JOURNAL			Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
PUBMED			Physical and Chemical Research (RIKEN), Laboratory for Genome
REFERENCE			Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
			RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama
			Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp,
			URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
			Fax: 81-45-503-9216]
			Please visit our web site (http://genome.gsc.riken.go.jp/) for
			further details.
			cDNA library was prepared and sequenced in Mouse Genome
			Encyclopedia Project of Genome Exploration Research Group in Riken
			Genomic Sciences Center and Genome Science Laboratory in RIKEN.
			Division of Experimental Animal Research in Riken contributed to
			prepare mouse tissues. First strand cDNA was primed with a primer
			[5'-GAGAGAGAGCGCGCAACTGAGTGTGTTTTTTTTTTTNN 3'], cDNA was
			prepared by using trehalose thermo-activated reverse transcriptase

FEATURES									
source									
Location/Qualifiers									
1. .1132									
/organism="Mus musculus"									
/mol_type="mRNA"									
/strain="C57BL/6J"									
/db_xref="PANTOM.DB:1110029M05"									
/db_xref="MGI:1892527"									
/db_xref="taxon:10090"									
/clone="1110029M05"									
/tissue_type="whole body"									
/clone_lib="RIKEN full-length enriched mouse cDNA library"									
/dev_stage="16-day embryo"									
15. .355									
/note="putative"									
ring_finger_protein 7 (MGD MGI:1337096)"									
/db_xref="MGI:1337096"									
1113. .1118									
/note="putative"									
1132									
polyA_site									
/note="putative"									
BASE COUNT 298 a 224 c 286 g 324 t									
ORIGIN									
Query Match 93.8%; Score 1069.4; DB 11; length 1132;									
Best Local Similarity 99.3%; Pred. No. 1.4e-218;									
Matches 1127; Conservative 0; Mismatches 1; Indels 7; Gaps 5									
QY	4	CTGCGCGCGCCCATGCGCCGACGTGAGAGACGCGCGAGGAAACCTGCGCTCTTCTTTCGCA	63						
Db	2	CTGCGCGCGCGCCCATGCGCCGACGTGAGAGAGCGCGAGGAAACCTGCGCTCTTCTTTCGCA	61						
QY	64	CTCCGGGAGACGAGGCTCCAAAGTCGGGAGGGGACAAAGATTTCTCTCAAGAGTGGAA	123						
Db	62	CTCCGGGAGACGCA-GCTCCAAAGTCGGGAGGGGACAAAGATTTCTCTCAAGAGTGGAA	120						
QY	124	CGCGGTAGCCATGTGGAAGCTGGGACGTTGAGTGCATACCTGTGCCATCTGCAGGGTCCA	183						
Db	121	CGCGGTAGCCATGTGGAAGCTGGGACGTTGAGTGCATACCTGTGCCATCTGCAGGGTCCA	180						
QY	184	GGGATGAGATGCGTCGCCCTTCGATGTCAAGCTGAAGAAACAACCAAGAGACCTGTGTGGT	243						
Db	181	GGGATGAGATGCGTCGCCCTTCGATGTCAAGCTGAAGAAACAACCAAGAGACCTGTGTGGT	240						
QY	244	CTGGGAGAGTGTAAACATTCCTTCCAACATGCTGCATGCTCCCTGTGGGTGAACAGAA	303						
Db	241	CTGGGAGAGTGTAAACATTCCTTCCAACATGCTGCATGCTCCCTGTGGGTGAACAGAA	300						
QY	304	CAATGCTGCGCTCTGTGCCACAGACGTGGGTATGTCAAAGATCGGCAATGAGAGGT	363						
Db	301	CAATGCTGCGCTCTGTGCCACAGACGTGGGTATGTCAAAGATCGGCAATGAGAGGT	360						
QY	364	GGGCCAGAGGCGCTCGGT	423						
Db	361	GGGCCAGAGGCGCTCGGT	420						
QY	424	TCATCCTTGAGAGAGAGAGATGCTGTGCGCTTGTGAAGCTCAACAAAGGCTTCTTAT	483						
Db	421	TCATCCTTGAGAGAGAGAGATGCTGTGCGCTTGTGAAGCTCAACAAAGGCTTCTTAT	480						
QY	484	TAATTTGTCTGTTTAGTTTGGGAAATTTCTCTAATAATTAAGATTAATTTGTTAAATGGC	543						
Db	481	TAATTTGTCTGTTTAGTTTGGGAAATTTCTCTAATAATTAAGATTAATTTGTTAAATGGC	540						
QY	544	CTTTCTCTACCTCTGGT	603						
Db	541	CTTTCTCTACCTCTGGT	600						
QY	604	AATGATCTTTGTTTATCTGTACCAAGACT--GGAACATTTGTTCACAAAGAAATTT	660						



Db		601	AATGATCTTTTGGTATTTCGTACCCAGCACTGTGGACAATTGTGTACAGAGAACAATT	660
Oy		661	GTTTGTTGTTTAGCTTGAGGGGTTAAAAAATAGATAAACGAATGTTACAGTAACAATAAA	720
Db		661	GTTTGTTGTTTAGCTTGAGGGGTTAAAAAATAGATAAACGAATGTTACAGTAACAATAAA	720
Oy		721	ATGCATTGAAAAAGCGACTCCTCTTAATCCCTTTTGTGTGGAGAGAGGCGAACGAGGC	780
Db		721	ATGCATTGAAAAAGCGACTCCTCTTAATCCCTTTTGTGTGGAGAGAGGCGAACGAGGC	780
Oy		781	CACCCTGCTGCTTCATTGCTTGCTGTAATGAGAGATTTAACTCGCACTCAGTAGAAGAGCG	840
Db		781	CACCCTGCTGCTTCATTGCTTGCTGTAATGAGAGATTTAACTCGCACTCAGTAGAAGAGCG	840
Oy		841	TAACTGTCGGGTTAAACCTGTAAATATGCGGTAACTGTCCGGGTAAACGGCTTTGTCTCTGAC	900
Db		841	TAACTGTCGGGTTAAACCTGTAAATATGCGGTAACTGTCCGGGTAAACGGCTTTGTCTCTGAC	900
Oy		901	TTCTCCCATCTTTGACTTTGGCCAGAGAGCGCTGATTTGCCAACCACTTAGTCTTAAAGAC	960
Db		901	TTCTCCCATCTTTGACTTTGGCCAGAGAGCGCTGATTTGCCAACCACTTAGTCTTAAAGAC	960
Oy		961	TGTTTTCTGTTTTGGCCAGAGGTGTATGTATGTTTTAGCAAATAATATAGAGAAA	1020
Db		961	TGTTTTCTGTTTTGGCCAGAGGTGTATGTATGTTTTAGCAAATAATATAGAGAAA	1018
Oy		1021	ATGGCTTACTAGTATTAACACTGAAAGTTTCATTATGCAATGTTTAAATAATATTTGCTT	1080
Db		1019	AT-GGTTACTAGTATTAACACTGAAAGTTTCATTATGCAATGTTTAAATAATATTTGCTT	1077
Oy		1081	TGAGTTATTAAAGTTTGATATATATCTGTTAAATCATTAATCAATTCATCAATT	1135
Db		1078	TGAGTTATTAAAGTTTGATATATATCTGTTAAATCATTAATCAATTCATCAATT	1132
RESULT 2				
AKO11328				
LOCUS				
DEFINITION		AKO11328	1114 bp mRNA linear HTC 05-DEC-2002	
		Mus musculus 10 days embryo whole body cDNA, RIKEN full-length		
		enriched library, clone:261005N22 product:ring finger protein 7,		
		full insert sequence.		
ACCESSION		AKO11328		
VERSION		AKO11328.1	GI:12847380	
KEYWORDS		HTC; CAP trapper.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE				
AUTHORS		1		
TITLE		Carninci, P. and Hayashizaki, Y.		
JOURNAL		High-efficiency full-length cDNA cloning		
MEDLINE		Meth. Enzymol. 303, 19-44 (1999)		
PUBMED		99279253		
REFERENCE		10349636		
AUTHORS		2		
TITLE		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
JOURNAL		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
MEDLINE		Normalization and subtraction of cap-trapper-selected cDNAs to		
PUBMED		prepare full-length cDNA libraries for rapid discovery of new genes		
REFERENCE		Genome Res. 10 (10), 1617-1630 (2000)		
TITLE		3		
JOURNAL		Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,		
MEDLINE		Kono, H., Akiyama, U., Nishi, K., Kitsumai, T., Taehiro, H., Itoh, M.,		
PUBMED		Sunai, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,		
REFERENCE		Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K.,		
AUTHORS		Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M.,		
TITLE		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,		
JOURNAL		Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.		
MEDLINE		RIKEN integrated sequence analysis (RISA) system--384-format		
PUBMED		Sequencing pipeline with 384 multiplexillary sequencer		
REFERENCE		11042159		

JOURNAL MEDLINE PUBMED REFERENCE	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4
AUTHORS	Kawai,J., Shitagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ichii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,Y., Fukuoka,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadoe,K., Matsuda,H., Ashburner,M., Batilova,S., Casavola,T., Fleischmann,W., Gaasterland,T., Gissi,C., Kling,B., Kochwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Peeble,G., Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M., Wagner,L., Mashio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarini,R., Barsh,G., Blake,Y., Boftelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Balt,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustinchik,S., Hill,D., Hotmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Monbaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Saeki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toy-o-Oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L., Wyshaw-Sorls-A., Yoshida,K., Hasegawa,Y., Kawai,J., Hongo,T., and Hayashizaki,Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length CDNA's Nature 420, 563-573 (2002) 6 (bases 1 to 1114)
TITLE	
JOURNAL	Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Funano,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kuwahara,C., Matsumura,T., Miyazaki,A., Nishii,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shibata,Y., Shitagawa,A., Shiraki,T., Sogabe,Y., Suuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toy-a,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
COMMENT	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, 1-7-22 Suicho-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, url:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGATCGACGATCCAGACTCTTTTCTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGACGTATTAAATTAATGCCCCCCCCC 3']. cDNA was cleaved with XhoI and SclI. Cloning sites, 5' end: XhoI, 3' end: SclI. Host: SOLR.
FEATURES	location/Qualifiers
SOURCE	1..1114 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J"

misc_feature		BASE COUNT		ORIGIN	
/db_xref="PANTOM DB:2610005N2" /db_xref="MG1:1893659" /db_xref="taxon:10090" /clone="2610005N2" /tissue_type="whole body" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev stage="10 days embryo" 15.._354 /note="putative ring finger protein 7 (MGI:1337096)" /db_xref="MG1:1337096"		289 a	220 c	287 g	318 t
Query	Match	91.3%	Score 1041;	DB 11;	Length 1114;
Best Local Similarity	99.3%;	Pred. No. 1.7e-212;			
Matches 1109;	Conservative 0;	Mismatches 0;	Indels 8;	Gaps 6	
Db	4	CTGCAGCCGCCCATGGCCGAGCGTGGAGAGACGCGAGAACCCCTGCGTCTTTCTTCGCA	63		
Qy	2	CTGCAGCCGCCCATGGCCGAGCGTGGAGAGACGCGAGAACCCCTGCGTCTTTCTTCGCA	60		
Db	64	CTCCGGGAGCGGAGGCTCCAACTCGGGAGGCGCAAGAAGTCTCTCTCAAGAATGGAA	123		
Qy	61	CTCCGGGAGCGGAGGCTCCAACTCGGGAGGCGCAAGAAGTCTCTCTCAAGAATGGAA	119		
Db	124	CGCGGTAGCCATGTGGAGCTGGGAGCGTTGAGTGCATACCTGTGCCATTCGAGGGTCCA	183		
Qy	120	CGCGGTAGCCATGTGGAGCTGGGAGCGTTGAGTGCATACCTGTGCCATTCGAGGGTCCA	179		
Db	184	GGTGAATGATGCTCTGCTTCGATATGTCAGCTGAATAAACAAGCAAGAGACTGTGTGGT	243		
Qy	180	GGTGAATGATGCTCTGCTTCGATATGTCAGCTGAATAAACAAGCAAGAGACTGTGTGGT	239		
Db	244	CTGGGAGAGGTAAACCATTCCTCCACAACACTGCTGCATGCTCCCTGTGGTGAAACAGA	303		
Qy	240	CTGGGAGAGGTAAACCATTCCTCCACAACACTGCTGCATGCTCCCTGTGGTGAAACAGA	299		
Db	304	CAATGCTGCTCTCTGTGCGACAGAGACTGGGGTAGTCCAAAGATGCGCAATGAGAGGT	363		
Qy	300	CAATGCTGCTCTCTGTGCGACAGAGACTGGGGTAGTCCAAAGATGCGCAATGAGAGGT	359		
Db	364	GGCCCAAGGGGCTCCGGTGTGGTGTGACCTCGGCAAAAGACTAAACCTGAGGGAGAT	423		
Qy	360	GGCCCAAGGGGCTCCGGTGTGGTGTGACCTCGGCAAAAGACTAAACCTGAGGGAGAT	419		
Db	424	TCATCCTTGAAGAGAGAGATGCTGTGCGCCCTTTGAGACTCCAAAGCTTCTTAT	483		
Qy	420	TCATCCTTGAAGAGAGAGATGCTGTGCGCCCTTTGAGACTCCAAAGCTTCTTAT	479		
Db	484	TAATTTGCTGTATTACTTTTGGAAATTTCTTCAATTTAAGAAATTTGTTAAATATGCG	543		
Qy	480	TAATTTGCTGTATTACTTTTGGAAATTTCTTCAATTTAAGAAATTTGTTAAATATGCG	539		
Db	544	CTTTCCCTACCTCGAGTGTGTGTGATACGATCATAGAAGACGGAACACACAGA	603		
Qy	540	CTTTCCCTACCTCGAGTGTGTGTGATACGATCATAGAAGACGGAACACACAGA	599		
Db	604	AATGATCTTTTATCTGTATCCACGACT--GGAACATGTGTTCAAGAAACATT	660		
Qy	600	AATGATCTTTTATCTGTATCCACGACTGTGGGAAATTGTGTTCAAGAAACATT	659		
Db	661	GTTTGCTTATGCTTGAAGGTTAAAAATAGATAACGAATGTTACAGTAAACAATTA	720		
Qy	660	GTTTGCTTATGCTTGAAGGTTAAAAATAGATAACGAATGTTACAGTAAACAATTA	719		
Db	721	ATGCAATGAAAGCCGACTCTCTTAATCTTTTGTGGAGAGAGGCAACGAGC	780		
Qy	720	ATGCAATGAAAGCCGACTCTCTTAATCTTTTGTGGAGAGAGGCAACGAGC	779		
Db	781	CACCCGTGCTTCAATTTGCTGTGAATGAGATTTTAACCTCAGTGAAGAGCG	840		
Qy	780	CACCCGTGCTTCAATTTGCTGTGAATGAGATTTTAACCTCAGTGAAGAGCG	839		

QY	841	TAAGTCGCGGGTAAACGTAAATATATGAGGTACGTGCGGTAAACCGCTTTGTCCTCGAC	900
Db	840	TAACGTGCGGGTAAACGTATATATATATGCGTACACTGCGGTAAACCGCTTTGTCCTCGAC	899
QY	901	TTCTCCATCTTTGACTTGCGCCAGGAACCGCTGATTTGTTCAACCACTTAGTTCTTAAAGAAC	960
QY	961	TGTTTTCTGTTTTTGGCCGAAGGTTGATTTGATGTTTCTTACGCAAAATATATAGAGGAAA	1020
Db	960	TGTTTTCTGTTTTTGG-CGAAGGTGATTTGATGTTTTAGTC-AAAATATATAGAGGAAA	1017
QY	1021	ATGGCTACTAGTAAACACTGAAGTTCATTATGCAATGTTTTATATATATATTTGTCCTT	1080
Db	1018	AT-GCTTACATGATTAACACGCAAGTTCATTATGCAATGTTTTATATATATATTTGTCCTT	1076
QY	1081	TGAGTTATTAAGTTTATATATATATATCTTTAAATCAT	1117
Db	1077	TGAGTTATTAAGTTTATATATATATATCTTTAAATCAT	1113
RESULT 3			
AK003248		1084 bp	linear
LOCUS			
DEFINITION			
AK003248			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			



Oy	901	TTTCGCATCTTGAATTCGCGCAGGAAGCTGGATTTGTCACCACTTTCGTAAGAAGC	960
Db	902	TTTCGCATCTTGAATTCGCGCAGGAAGCTGGATTTGTCACCACTTTCGTAAGAAGC	961
Oy	961	TGTTTTCTGTTTTTGGCGGAAGGTGGTATTGTATGTTTTTATGCAAAAATATATAGTGAA	1020
Db	962	TGTTTTCTGTTTTTGGCGGAAGGTGGTATTGTATGTTTTTATGCAAAAATATATAGTGAA	1019
Oy	1021	ATGGTCTCTGTATTAACAAGCACTTCATTAATGAAGTCTTTAATAAATATAGTGCT	1080
Db	1020	AT-GCTTACTGATTAACAAGCACTTCATTAATGAAGTCTTTAATAAATATAGTGCT	1078
Oy	1081	TGAGTT 1086	
Db	1079	TGAGCT 1084	
RESULT 4			
LOCUS	AK002628	1078 bp mRNA linear HTC 05-DEC-2002	
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610013021 product:ring finger protein 7, full insert sequence.		
ACCESSION	AK002628		
VERSION	AK002628.1 GI:12832752		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
TITLE	Genome Res. 10 (10), 1617-1630 (2000)		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	2049374		
PUBMED	11042159		
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Teshio, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
AUTHORS	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
TITLE	Genome Res. 10 (11), 1757-1771 (2000)		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flieschman, M., Gaasterland, T., Gissi, C., King, B., Kochia, H., Kuell, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, P., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carinini, P., de Bonaldo, M. F., Brownstein, M. J., Bul, C., Flecher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, K., Lee, N. H., Lyons, P., Marchionni, L., Mashima, Y., Mazzarelli, J., Mombaerts, P., Nordone, P.,		

TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	
AUTHORS	5
TITLE	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 1078)
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	Aachai,J., Aizawa,K., Akhira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirakawa,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kikukawa,T., Kato,H., Kawai,J., Kojima,Y., Komno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL: <a href="http://genome.gscc.riken.go.jp/">http://genome.gscc.riken.go.jp/</a> , Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gscc.riken.go.jp/">http://genome.gscc.riken.go.jp/</a> ) for further details
TITLE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCGCACTCGAGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rct = 5.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGAGATCCAGAGCTCAATTATTAATTAAACCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES	Location/Qualifiers
SOURCE	1..1078
FEATURES	/organism="Mus musculus"
SOURCE	/mol_type="mRNA"
FEATURES	/strain="C57BL/6J"
FEATURES	/db_xref="FANTOM DB:0610013021"
FEATURES	/db_xref="MG1:1892044"
FEATURES	/db_xref="taxon:10090"
FEATURES	/clone="0610013021"
FEATURES	/sex="male"
FEATURES	/tissue="kidney"
FEATURES	/clone_lib="RIKEN full-length enriched mouse cDNA library"
FEATURES	/dev_stage="adult"
FEATURES	15..354
FEATURES	/note="putative
FEATURES	ring finger protein 7 (MGD MG1:1337096)"
FEATURES	1060..1065
FEATURES	/note="putative"
FEATURES	1078
FEATURES	/note="putative"
BASE COUNT	276 a 216 c 283 g 303 t



Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedic: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

source 1. 997  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1110029M05"  
 /tissue\_type="whole body"  
 /dev\_stage="18-day embryo"  
 /clone\_lib="RIKEN full-length enriched, 18-day embryo whole body"

BASE COUNT 250 a 206 c 270 g 271 t

## ORIGIN

### Query Match

Best Local Similarity 98.8%; Score 875; DB 14; Length 997;

Matches 924; Conservative 0; Mismatches 5; Indels 6; Gaps 4;

QY 11 GCCGCATGCGCCGAGTGGAGGACGCGAGAACCTCGTCTTCTTCGACCTCCGG 70  
 Db 1 GCCGCATGCGCCGAGTGGAGGACGCGAGAACCTCGTCTTCTTCGACCTCCGG 60  
 QY 71 AGCGGAGGCTCCAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGGAAACCGGTA 130  
 Db 61 AGCGGA-GCTCCAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGGAAACCGGTA 119  
 QY 131 GCCATGTGAGTGGGAGGCTGATAGTGCATCTGCGCATCGAGGCTCCAGGTGATG 190  
 Db 120 GCCATGTGAGTGGGAGGCTGATAGTGCATCTGCGCATCGAGGCTCCAGGTGATG 179  
 QY 191 GATGCTGCTTGCATGTCAAGCTGCAAAACAAGAGACTGTGTTGTTGGTGGGA 250  
 Db 180 GATGCTGCTTGCATGTCAAGCTGCAAAACAAGAGACTGTGTTGTTGGTGGGA 239  
 QY 251 GAGTGAACCATTTCTTCCACAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310  
 Db 240 GAGTGAACCATTTCTTCCACAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299  
 QY 311 TSCCCTCTGTCAGAGGAGTGGGTAGTCCAAAGATGCGCAATGAGAGTGGCCG 370  
 Db 300 TSCCCTCTGTCAGAGGAGTGGGTAGTCCAAAGATGCGCAATGAGAGTGGCCG 359  
 QY 371 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430  
 Db 360 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419  
 QY 431 TGAAGAGAGAGAGAGTGTGGGCTTTGAGACTCAACCAAGGCTTTGCTTTATTTG 490  
 Db 420 TGAAGAGAGAGAGAGTGTGGGCTTTGAGACTCAACCAAGGCTTTGCTTTATTTG 479  
 QY 491 TCTGTTAGTTTGGGAAATTTCTCAATTAAGATTAATTTGTTAAATGSCCTTTCT 550  
 Db 480 TCTGTTAGTTTGGGAAATTTCTCAATTAAGATTAATTTGTTAAATGSCCTTTCT 539  
 QY 551 ACCCTGTGTGTGTGTGTGATAGCATAGAGAGAGCAACCCAGAAATGATC 610  
 Db 540 ACCCTGTGTGTGTGTGTGATAGCATAGAGAGAGCAACCCAGAAATGATC 599

QY 611 TTGTTATCTGTAACCAACGAC--TGAACATGTTGTTCAACAAGAACATGTTGTG 667  
 Db 600 TTTGTTATCTGTAACCAACGACTGTGGACATGTTGTTCAACAAGAACATGTTGTG 659  
 QY 668 TTTATGCTGAGGTTAAAAATAGATTAAGATTTAGAGTAAACAATAATGACTT 727  
 Db 660 TTTATGCTGAGGTTAAAAATAGATTAAGATTTAGAGTAAACAATAATGACTT 719  
 QY 728 GAAAGCGCACTCTCTTAATCTCTTTTGTGTGGAGAGAGGAGGAGGCAACCTG 787  
 Db 720 GAAAGCGCACTCTCTTAATCTCTTTTGTGTGGAGAGAGGAGGAGGCAACCTG 779  
 QY 788 CTGCTTCAATTTGCTGTAATGAGATTTTAACTGCACTAGTGAAGAGCGTACTGT 847  
 Db 780 CTGCTTCAATTTGCTGTAATGAGATTTTAA-CTGCACTAGTGAAGAGCGTAACTGT 838  
 QY 848 CGGGTAACTGTAAATATGCGGTAATCTGTGCGGTTAAAGCGCTTTGCTGCTGCTTCCA 907  
 Db 839 CGGGTAACTGTAAATATGCGGTAATCTGTGCGGTTAAAGCGCTTTGCTGCTGCTTCCA 898  
 QY 908 TCTTTGACTTGGCCAGGAGAGCGCTGATTTGTTCAAC 942  
 Db 899 TCTTTGACTTGGCCA-GAAGGCTGATTTGTTCAAC 932

## RESULT 6

BY703542

LOCUS

## DEFINITION

BY703542 RIKEN full-length enriched, 18-day embryo whole body Mus

musculus cDNA clone 1110001M08 5', mRNA sequence.

## VERSION

BY703542.1 GI:27114651

## KEYWORDS

EST.

## ORGANISM

Mus musculus (house mouse)

## SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

AUTHORS

1 (bases 1 to 988)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nishida, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matuda, H., Batalov, S.,  
 Betsel, K.W., Blake, J.A., Brad, D., Brusc, V., Chochua, C., Corbett,  
 L.E., Cousins, S., Dalla, E., Dregan, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pereira, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,  
 Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 R.Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C.A., Setou,  
 M., Shimada, K., Sultana, R., Takenawa, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wysshaw-Boriss, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komori, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Heshizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Watanabe, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

## TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

2234683

1246851

1246851

## COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute





```

/db_xref="taxon:10090"
/clone="IMAGE:6518791"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 134"
/Note="Vector: pCMV-Sport6.1.ccd; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by Resgen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT      255 a      187 c      246 g      259 t      2 others
ORIGIN

Query Match      73.4%; Score 836.2; DB 13; Length 949;
Best Local Similarity 96.2%; Pred. No. 1.2e-168;
Matches 910; Conservative 0; Mismatches 29; Indels 7; Gaps 5;

QY 82 CAAGTCGGAGGCGCAAGATTTCTCTCAAGAAAGTGAAGCGGGTAGCCATGTGAG 141
DB 4 CAAGTCGGAGGCGCAAGATTTCTCTCAAGAAAGTGAAGCGGGTAGCCATGTGAG 63
QY 142 CTGGGACGTTGAGTGCATCTGTCATCTGACAGGCTCAAGTATGATGCTGCT 201
DB 64 CTGGGACGTTGAGTGCATCTGTCATCTGACAGGCTCAAGTATGATGCTGCT 123
QY 202 TCGATGTCAAGCTGAAAAAACAAGACAGAGACGTGTGTGTGTGGGAGATGTAACA 261
DB 124 TCGATGTCAAGCTGAAAAAACAAGACAGAGACGTGTGTGTGTGGGAGATGTAACA 183
QY 262 TTCCTTCCACACTGTGATGTCCTGCGGGGTAAGAAACAATCGCTGCTCTGTG 321
DB 184 TTCCTTCCACACTGTGATGTCCTGCGGGGTAAGAAACAATCGCTGCTCTGTG 243
QY 322 CCAGCAGAGACTGGTGTGTCACAAAGATCGCAATATGAGAGTGGCCAGCGCTCTGT 381
DB 244 CCAGCAGAGACTGGTGTGTCACAAAGATCGCAATATGAGAGTGGCCAGCGCTCTGT 303
QY 382 GTGTGTGCTGACCTCTGACCAAGACCTAAGACCTGACGGGGATTCATCTTGAAGAGAGA 441
DB 304 GTGTGTGCTGACCTCTGACCAAGACCTAAGACCTGACGGGGATTCATCTTGAAGAGAGA 363
QY 442 GGATGCTGTGCGCTTTGAGACCTCAACCAAGGCTTCTTATTTATTTGTCTGTAGTT 501
DB 364 GGATGCTGTGCGCTTTGAGACCTCAACCAAGGCTTCTTATTTATTTGTCTGTAGTT 423
QY 502 TTGGGAAATCTCTACAAATTAAGATAATTTGTTAAAAATGCGCTTCTTACCTGTGTGT 561
DB 424 TTGGGAAATCTCTACAAATTAAGATAATTTGTTAAAAATGCGCTTCTTACCTGTGTGT 483
QY 562 GTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
DB 484 GTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 622 GTACCCAGCACT--GGAACATGTGTTCAGAGAAACATTTGTTTATGCTTGA 678
DB 544 GTACCCAGCACTGTGGGAACATTTGTTTCAGAGAAACATTTGTTTATGCTTGA 603
QY 679 GGGTTAAAAAATGATTAAGAAATGTTAAGTAACAATTAATGATGAAAGCCGAC 738
DB 604 GGGTTAAAAAATGATTAAGAAATGTTAAGTAACAATTAATGATGAAAGCCGAC 663
QY 739 TCTCTCTAATCC--TTTTGTTGGGAGAGAGCGAGCGACCTGCTGTCTTCA 797
DB 664 TCTCTCTAATCC--TTTTGTTGGGAGAGAGCGAGCGACCTGCTGTCTTCA 723
QY 798 TTGCTGTGATGAGATTTTAACTGCACTCACTGAGAGAGCGTAACTGTGGGTAAC 857
DB 724 TTGCTGTGAGAAATGAGATTTTAACTGCACTCACTGAGAGAGCGTAACTGTGGGTAAC 783
QY 858 GTTAATATGGGTAACGTGGGT--AAAGCGCTTTGTCTCTGACTTCTCATCTTTGACT 916
DB 784 GTTAATATGGGTAACGTGGGTAAACGGCTTTGTCTCTGACTTCTCATCTTTGACT 843
QY 917 TTGCAAGG--AAGCTGGAATGTTCAACCACTTAGTTCTAAGAACTGTTTCTGTTTGG 975

```

```

DB 844 TTGGCAGAGAAACCTGGAATGTTCAACCACTAGTTTAAAGACNTTTTCCGGTTTG 903
QY 976 CCGAAGGTTGATTTGAT--GTTTATGTCAAAAATATTATGAGGAAA 1020
DB 904 CCAAGGTTGATGATGATGATTTTATGAGCCCAAAATATTAGTTAGAAA 949

RESULT 8
CB204521
LOCUS
DEFINITION
AGENCOCUR 11276019 NIH MGC 135 Mus musculus cDNA clone
IMAGE:30138707 5', mRNA sequence.
CB204521
VERSION
CB204521.1 GI:28242067
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 901)
NIH-MGC http://mgi.nci.nih.gov/
JOURNAL
Unpublished
COMMENT
Contact: Robert Strauberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: NDA0041 row: p column: 12
High quality sequence stop: 678.
Location/Qualifiers
1..901
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30138707"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 135"
/Note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized full-length enriched library from pooled mouse
embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5
and 15.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.6k bp.
Normalization (Cot value): 7.5 kb. Priming sequence:
5'GACTAGTCTTAATGCGCAGCGCGCCGCT(7)3' Tissue contributed by
David Rowe. Library constructed by Resgen, Invitrogen
Corp."
BASE COUNT      231 a      190 c      241 g      239 t
ORIGIN

Query Match      73.1%; Score 833.6; DB 14; Length 901;
Best Local Similarity 98.3%; Pred. No. 4.4e-168;
Matches 885; Conservative 0; Mismatches 9; Indels 6; Gaps 4;

QY 21 CCGAGTGTGAGAGAGCGGAGAGAACCTGCTCTTTCTTGCACATCCGAGAGCGAGGCT 80
DB 1 CCGAGTGTGAGAGAGCGGAGAGAACCTGCTCTTTCTTGCACATCCGAGAGCGAGGCT 60
QY 81 CCAAGTCGGAGAGCGGAGCAAGATTTCTCTCAAGAAATGGAAGCGGGTATGATGTA 140
DB 61 CCAAGTCGGAGAGCGGAGCAAGATTTCTCTCAAGAAATGGAAGCGGGTATGATGTA 120
QY 141 GCTGGGAGGTTGAGAGGATACCTGTGCACTCTGAGAGGTCAGAGTGAATGATGCTGCC 200
DB 121 GCTGGGAGGTTGAGAGGATACCTGTGCACTCTGAGAGGTCAGAGTGAATGATGCTGCC 180
QY 201 TTGATGTCAAGCTGAAAAAACAAGACAGAGACGTGTGTGTGTGTGGGAGAGTGAACC 260

```



```

Db      181 TTGGATGTCAAGCTGAAAAACAAGACAGACGTGTGTGGTGGGAGAGTGAAC 240
Qy      261 ATTCTTCCCACTGCTGATGTCCCTGGGTGAACAACATCGTCCCTCTGT 320
Db      241 ATTCTTCCCACTGCTGATGTCCCTGGGTGAACAACATCGTCCCTCTGT 300
Qy      321 GCCAGAGAGCTGGGTAGTCCAAAGATGGCAATGAGAGTGGCCAGGCGCTCTGT 380
Db      301 GCCAGAGAGCTGGGTAGTCCAAAGATGGCAATGAGAGTGGCCAGGCGCTCTGT 360
Qy      381 TGTGTGTGTGACCTGACCAAGACTAAACACTGAGGGGATTCATCTTGAAGAGAG 440
Db      361 TGTGTGTGTGACCTGACCAAGACTAAACACTGAGGGGATTCATCTTGAAGAGAG 420
Qy      441 AGGATGCTGTGGCCCTTTGAGACTCACCAAGGCTGCTTTAATTAATTTGTCTTTAGT 500
Db      421 AGGATGCTGTGGCCCTTTGAGACTCACCAAGGCTGCTTTAATTAATTTGTCTTTAGT 480
Qy      501 TTGGGAAATTCCTCAATTAAGATTAATTTGTTAAATAGGCTTCTTCTACTGTGTG 560
Db      481 TTGGGAAATTCCTCAATTAAGATTAATTTGTTAAATAGGCTTCTTCTACTGTGTG 540
Qy      561 TGTGTGTGTGATCGAATGATGAAGAGGAGAACACCAAGAAATGATCTTTGTTATC 620
Db      541 TGTGTGTGTGATCGAATGATGAAGAGGAGAACACCAAGAAATGATCTTTGTTATC 600
Qy      621 TGTACCCAGACCT--GGAACATTTGTGTACAGAAAGACATTTGTTGTATATGCTTG 677
Db      601 TGTACCCAGACCTGTGGGAACATTTGTGTACAGAAAGACATTTGTTGTATATGCTTG 660
Qy      678 AGGGTAAATAATAGATTAACGAATGTTTACGTAACAAATGAATGATTAAGAAAGCGA 737
Db      661 AGGGTAAATAATAGATTAACGAATGTTTACGTAACAAATGAATGATTAAGAAAGCGA 720
Qy      738 CTCCTCT--AATCCTTTTGTGTGGAGAGAGGCAAGCGCC--ACCCTGCTGTCTTC 795
Db      721 CTCCTCTAATCTTTTGTGTGGAGAGAGGCAAGCGCC--ACCCTGCTGTCTTC 780
Qy      796 ATTGTCTGATGATGAGATTTTAACCTGCACTGATGAAGAGCGGTAAGTGTGGGTAA 855
Db      781 ATTGTCTGATGATGAGATTTTAACCTGCACTGATGAAGAGCGGTAAGTGTGGGTAA 840
Qy      856 CTGTATATAGGCGGTAACTGT--GGGTAAAGGCTTTGCTCCGACTCTCCATCTTTGA 914
Db      841 CTGTATATAGGCGGTAACTGTCCGGTTAAAGCTTTTGTCTCCGACTCTCTCATCTTGA 900

RESULT 9
LOCUS   B0964159 1185 bp mRNA linear EST 21-NOV-2002
DEFINITION
ACCESSION B0964159
VERSION   B0964159
KEYWORDS  IMAGE:6509718 5', mRNA sequence.
SOURCE    B0964159.1 GI:22379637
ORGANISM  Mus musculus (house mouse)
MUSCULUS  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1185)
NIH-WGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fda.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

```

```

FEATURES
source
    Plate: LLM14077 row: h column: 07
    High quality sequence srop: 704.
    Location/Qualifiers
        1..1185
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:6509718"
            /issue_type="undifferentiated limb"
            /lab_host="DH10B (phage-resistant)"
            /clone_11b="NIH_MGC_134"
            /note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
            NotI; Cloned unidirectionally. Primer: Oligo dt. Average
            insert size 1.7 kb. Constructed by Reegen, Invitrogen
            Corp. Note: this is a NIH_MGC Library."
BASE COUNT  286 a 256 c 331 g 307 t 5 others
ORIGIN
Query Match 70.4%; Score 802.4; DB 13; Length 1185;
Best Local Similarity 94.8%; Pred. No. 2,1e-161;
Matches 929; Conservative 0; Mismatches 33; Indels 18; Gaps 9;

Qy      1 GTTTCGCGCGCGCCAGTCCGACGCTGAGAGACGCGAGAAACCTGCGCTTCTTC 60
Db      66 GTTTCGCGCGCGCCAGTCCGACGCTGAGAGACGCGAGAAACCTGCGCTTCTTC 125
Qy      61 GCACTCCGGAGACCGAGGCTCCAGTCGCGAGAGCGCAAGATTTCTCTCAAGAGTC 120
Db      126 GCACTCCGGAGACCGAGGCTCCAGTCGCGAGAGCGCAAGATTTCTCTCAAGAGTC 185
Qy      121 GAACGCGGTAGCATGTGAGCTGGGACGTTGATGTCGATACCTGTCAGAGGGT 180
Db      186 GAACGCGGTAGCATGTGAGCTGGGACGTTGATGTCGATACCTGTCAGAGGGT 245
Qy      181 CCAAGTATGATGATGCTGCTTGTGATGTCAGCTGAACCTGAACCAAGACAGACTGTGTG 240
Db      246 CCAAGTATGATGATGCTGCTTGTGATGTCAGCTGAACCTGAACCAAGACAGACTGTGTG 305
Qy      241 GGTCTGGGAGAGTGTAACTTCTTCCAACTGCTGATGTCGATGTCGCTGTGGTGAACA 300
Db      306 GGTCTGGGAGAGTGTAACTTCTTCCAACTGCTGATGTCGATGTCGCTGTGGTGAACA 365
Qy      301 GAACAATGCTGCGCTCTGTGCGACAGAGACTGGGTAGTCCAAAGATGGCAATGAGA 360
Db      366 GAACAATGCTGCGCTCTGTGCGACAGAGACTGGGTAGTCCAAAGATGGCAATGAGA 425
Qy      361 GGTGCGCAGGCGCTCTGTGCTGTGTGCTGACCTTGACCAAGACTTAACAAGTGAAGG 420
Db      426 GGTGCGCAGGCGCTCTGTGCTGTGTGCTGACCTTGACCAAGACTTAACAAGTGAAGG 485
Qy      421 GATTCATCTTGAAGAGAGAGATGCTGTGCGCTTTGAAGCTCACCAAGGCTTGTCTT 480
Db      486 GATTCATCTTGAAGAGAGAGATGCTGTGCGCTTTGAAGCTCACCAAGGCTTGTCTT 545
Qy      481 TATTAATTTGCTGTTTGTAGTTTGGGAAATCTCTACATTAAGATTAATTTGTTAAAT 540
Db      546 TATTAATTTGCTGTTTGTAGTTTGGGAAATCTCTACATTAAGATTAATTTGTTAAAT 605
Qy      541 GGCCTTCTTACCTGCTGTGTGTGTGTGATACGATCATGAGAGAGAGACCAACCA 600
Db      606 GGCCTTCTTACCTGCTGTGTGTGTGTGATACGATCATGAGAGAGAGACCAACCA 665
Qy      601 GAAATGATCTTTGTTTATCTGATCCAGCACT--GGAACATTTGTTTCAAGAAAGC 657
Db      666 GAAATGATCTTTGTTTATCTGATCCAGCACTTTGTTTCAAGAAAGC 725
Qy      658 ATTGTTGTTTATGCTTGAAGGTTAAATAATGATTAACGAATGTTACAGTAACAAT 717
Db      726 ATTGTTGTTTATGCTTGAAGGTTAAATAATGATTAACGAATGTTACAGTAACAAT 785
Qy      718 AAAATGATTAAGAAAGCGACTCTCTTAATCC--TTTGTGTGTGAGAGAGGCAAGC 775
Db      786 AAAATGATTAAGAAAGCGACTCTCTTAATCCNNTTTGTGTGTGAGAGAGGCAAGC 845

```

QY	776	GAGG-CCACCCGTCGTCTTCAATTGGCTGTG--AATGAGATTTTAACCTGACTCACTG	832
Db	846	GAGGCCCAACCCGTGCGNCTATTGGCTGGGGAATGANGGATTTTAACTGCACTCACTG	905
QY	833	AAG-AGGGGTACTGTGCGGTAAACTG---TAATATGGCGTAACGTG---GGGTAAACG	885
Db	906	AAGAGGCGCTACTGTTCGGGCTAACTGAAAAAATGGCGCTAACTGTTCGGGGTAAACGG	965
QY	886	GCTTTGTCTCTG-ACTTCTCATCTTTGACTTTGGCCA--GGAAGCCTGATTTGTTCAAC	942
Db	966	CTTTTGTCTCTGAACTTCTCATCTTTTGACTTTGGCCAAGAAAACTGATTTGGTTTCAAC	1025
QY	943	CACCTAGCTTCTAAGAACTG	962
Db	1026	CACCTAAGTCTTAAAGAACAG	1045

RESULT	10
BQ964115	
LOCUS	BQ964115 968 bp mRNA linear EST 21-AUG-2002
DEFINITION	AGENCOURT_10053663 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:	6514192 5' , mRNA sequence.

ACCESSION	B0964115	GI:22379593
VERSION	B0964115.1	
KEYWORDS	EST.	
SOURCE	Mus musculus	(house mouse)
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 968)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC).  
Unpublished  
Contact: Robert Strusberg, Ph.D.

Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://Image.lnl.gov>  
Plate: LLM14089 row: b column: 17  
High quality sequence stop: 566.

FEATURES	SOURCE
location/Qualifiers	1..968
/organism="Mus musculus"	
/mol_type="mRNA"	
/db_xref="taxon:10090"	
/clone="IMAGE:6514192"	
/tissue_type="undifferentiated limb"	
/lab_host="DH10B (phage-resistant)"	
/clone_lib="NIH MGC 134"	
/note="vector: pCMV-SPORT6.1.cdbb, Site_1: EcoRV, Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC Library."	
BASE COUNT	288 a 170 c 219 g 282 t 9 others

Origin	
Query Match	68.6%; Score 781.8; DB 13; Length 968;
Best Local Similarity	98.8%; Pred. No. 5.5e-15;
Matches 840: Conservative	0; Indels 3; Gaps 5;

293 GTGAACAGAACATCGTCGCTCTGTGCGACAGGACTGGTAGTCCAAAGATCGC 352  
1 GTGAACAGAACATCGTCGCTCTGTGCGACAGGACTGGTAGTCCAAAGATCGC 60

QY 353 AATGAGAGGTGGCCCAAGCGCTCTGTGTGGTTGCTGACCTTGGA CAAGA CTAACA 412  
61 AATGAGAGGTGGCCCAAGCGCTCTGTGTGGTTGCTGACCTTGGA CAAGA CTAACA 120

Oy	413	CTGAGGGGATTCATCTCTTAGAGAGAGAGATCTGTGGCCCTTTGAGACTACCCAAAG	472
Db	121	CTGAGGGGAGATCATCTCTTAGAGAGAGAGAGATCTGTGGCCCTTTGAGACTACCCAAAG	180
Oy	473	GCTTGGCTTATTAATTTGTCTGTTAAGTTTGGGAAATTCCTACAAATTAGATTAATTTG	532
Db	181	GCTTGGCTTATTAATTTGTCTGTTAAGTTTGGGAAATTCCTACAAATTAGATTAATTTG	240

**OY**            533 TTTAAATGGCCTTTCCTACCTGTGNGTGNGTGATACGAATGCATAGAAGCGA 592  
**Dd**            241 TTAAATGGCCTTCTACTCTGTGNGTGNGTGATACGAATGCATAGAAGCGA 300			

**OY**      593 GAAACCAAGAAAATGATCTTTGTATTCTGTACCAGACT---GGACAATTGTTCAC 649  
|||||  
**Dd**      301 GAACACCAAGAAAATGATCTTTGTATTCTGTACCAGACTGTGGGAACAATGTGTTCAC 360  
|||||

Qy 650 AGAAGACATTGTTTGTTATGCTTGAGGCTTAAAAATAGATTAACCAATGTACAG 709  
Db 361 AGAAGACATTGTTTGTTATGCTTGAGGCTTAAAAATAGATTAACCAATGTACAG 420

770	GCAGCGAAGGCGACCCGCGCTCTCTTCACTTTTGGCGCTGATATATGAGATTTTATATCGGCGACCTCA	839	
DY			
QY	710	TACCAATTAATAATGCAATGTAAAAACGCACTCCCTCTATCTCTTTGTGTGTGGAGAGAG	769
Db	421	TACCAATTAATAATGCAATGTAAAAACGCACTCCCTCTATCTCTTTGTGTGTGGAGAGAG	480

[illegible]

D5 541 GTAAAGAGGCGTAACTCTCGGGTAAACTGTAAATATGCGCTAACTGTTCGGGTAAACGGCTT 600

D6 890 TGTCTCTGACTTCTCCATCTTTGACCTGGCCAGAAACCTGGATTGTTCAACACACTTAA 949

Db	Qy
601	950
<p>           TGTCTCCTGACTTCCATCTTGCATCTTGACCTTGCCACGAGACCTGATTTGTTCAACCACTTNG            660         </p>	<p>           TTCTAAAGAACTGTTTTCTGTTTTTTCGCCGAGGTGTGATTGTATGTTTTAGTCAAAAAATA            1009         </p>

Db 661 TTCTAAGAACTGTTTTCGTCTTTTTC-CGAGGTTGATTTGATGTTTAAATC-AAAAA 718

Qy 1010 TTAGTAGAAATGGCTTACTACTATATACACGACGTCATTATGCATGTTTAAATAA 1069

Db	719	TTAGTAGGAAAT-GCTTACTAGCTAATACACCTGAAGTCATTGCAATGTTTTAATAAA	777
Qy	1070	ATAATGT-GCTTGAAGTTATTAAGTTGATATATACCTTAAATCAITTAACATAATC	1128
Db	778	AAATTCGGCTTTGACATTTTATATAGCTATATATCTCTTTAAATATATTAATCTATTC	837

QY	1129	ATCAATTAAA	1138
Db	838	ATCAATTAAA	847

```
RESULT 11  
BI732003  
  
*****  
  
*****  
  
*****
```

LOCUS	820 bp	mRNA	EST 20-SEP-2000
BI732003		linear	
DEFINITION	6033558656F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363056 5',		
	mRNA sequence.		
ACCESSION	BI732003		
VERSION	BI732003.1		
	GI:15709016		

**KEYWORDS** EST.  
**SOURCE** *Mus musculus* (house mouse)  
**ORGANISM** *Mus musculus*  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
1 (bases 1 to 820)  
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
AUTHORS  
TITLE  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
10.1016/j.ygeno.2004.08.005

UNPUBLISHED  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph. D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: The Cepco Laboratory  
cDNA Library Preparation: Life Technologies, Inc.



Db	67	GAGCGCACAAGATGTTCTTCTCTAAGAAGTGGAACGGGGTAGTCAGCATGTGGAC	126
Oy	149	GTTGAGTGCATACCTGTGCCATCTGCAGGGTCCAGGTGATGATGCTTGCATGT	208
Db	127	GTTGAGTGCATACCTGTGCCATCTGCAGGGTCCAGGTGATGATGCTTGCATGT	186
Oy	209	CAACTGAAAAAACAAGCAAAGAGAACTGTGTGTGTGGCTGGGGAGATGAACATTCCCTC	268
Db	187	CAGGTGAAAAACAAGCAAAGAGAACTGTGTGTGTGTGGAGAGATGAACATTCCCTC	246
Oy	269	CACAACCTGCTCATGTCCCCTGTGGGTGAACAGAACAAATGCTGCGCTCTGTGCACAG	328
Db	247	CACAACCTGCTCATGTCCCCTGTGGGTGAACAGAACAAATGCTGCGCTCTGTGCACAG	306
Oy	329	GACTGGGTATGTCCAAAGAAATGCGCAATAGAAGGTGGCCAGGGCTCCGTGTGTGTG	388
Db	307	GACTGGGTATGTCCAAAGAAATGCGCAATAGAAGGTGGCCAGGGCTCCGTGTGTGTG	366
Oy	389	CTGACCCTGGACAAGAACTTAACAATGTCAGGGGATTCATCCTTGAGAGAGAGATGCT	448
Db	367	CTGACCCTGGACAAGAACTTAACAATGTCAGGGGATTCATCCTTGAGAGAGAGATGCT	426
Oy	449	GTCGGCCTTTGAGACTCACCAAGGCTTGCTTATTAATTTGTCTGTTAAGTTTTGGAA	508
Db	427	GTCGGCCTTTGAGACTCACCAAGGCTTGCTTATTAATTTGTCTGTTAAGTTTTGGAA	486
Oy	509	AATTCCTCAATTAAGATAATTTGTTAAAATGCGCTTCTTCTACTCTGTAGTGTGTG	568
Db	487	AATTCCTCAATTAAGATAATTTGTTAAAATGCGCTTCTTCTACTCTGTAGTGTGTG	546
Oy	569	TGATACGAATGCATAGAAAGAGCGAACAACCAAGAAATGATCTTGTTTATCTGACCCA	628
Db	547	TGATACGAATGCATAGAAAGAGCGAACAACCAAGAAATGATCTTGTTTATCTGACCCA	606
Oy	629	CGACT--GGAACATGTGTTCAGAGAAACATGTTGTGTTATGCTGAGGGTTAA	685
Db	607	CGACTGTGGGAACATGTGTTCAGAGAAACATGTTGTGTTATGCTGAGGGTTAA	666
Oy	686	AAATAGATPAAACGAATGTTACAGTAAACAAT-AAATGCAATGTAAGAAACCGACTCC-TC	743
Db	667	AAATAGATPAAACGAATGTTACAGTAAACAAT-AAATGCAATGTAAGAAACCGACTCC-TC	726
Oy	744	CTAATCCTTTTGTGTGTGGAGAGAGGCAAGGAGGCCACCTGCTGTGC-TTCATTGCT	802
Db	727	CTAATCCTTTTGTGTGTGGAGAGAGGCAAGGAGGCCACCTGCTGTGTGCATTGCT	786
Oy	803	G--TGAATGAGGATTTTAACT-GCACTCAGTGAAG-AAGCGTAACTGCGGTTAACTG	858
Db	787	GGTGGAATGAGGATTTTAACTGCACTGCACTCAGTGAAGAGGCGTAACTGCGGTTAACTG	846
Oy	859	TAATATGCGTAACTGTGCGGTTAAACGGCTTTGTCTC	895
Db	847	TAATATGCGTAACTGTGCGGTTAAACGGCTTTGTCTC	883
RESULT_13			
CA980368		882 bp mRNA linear EST 06-JAN-2003	
LOCUS			
DEFINITION	AGENCOURT_11294967 NIH MGC_164 Mus musculus cDNA clone		
ACCESSION	CA980368		
VERSION	CA980368.1 GI:27513022		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 882)		
JOURNAL	NIH-MGC http://mgi.mcg.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Strausberg Ph.D.		

	Email:	cgabs-j@mail.nih.gov
	Tissue Procurement:	Dr. David Rowe and Dr. Mina
	CDNA Library Preparation:	Invitrogen Corp
	CDNA Library Arrayed by:	The I.M.A.G.E. Consortium (LNLN)
	DNA Sequencing by:	Agencourt Bioscience Corporation
	Clone distribution:	MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov plate: NDAM0059 row: 0 column: 18 High quality sequence stop: 662.
FEATURES	Location/Qualifiers	
source	1..882	/organism="Mus musculus"
	/mol_type="mRNA"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:30145601"	
	/lab_host="DH10B (phage-resistant)"	
	/clone_id="NH_MGC_164"	
	/note="Vector: pCMV-SPORT6.1.ccdB; Site_1: EcoRV; Site_2: NotI; Non-normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) Cloned directionally, priming method: Oligo-dT. cDNA enrichment: -lik bp. Average insert size 1.8k bp. Priming sequence: 5' GACTGCTTCATACATCGCAGCGGCCGCC(T) 3'. Tissue contributed by, David Rowe. Library constructed by ResGen Invitrogen Corp."	
BASE COUNT	234 a 173 c 230 g 244 t	1 others
ORIGIN		
Query Match	64.9%	Score 740; DB 14; Length 882;
Best Local Similarity	97.4%	Pred. No. 4.9e-148;
Matches 785; Conservative	0; Mismatches 15; Indels 6; Gaps 3	
OY	191	GATCCCTGCTTCATGTCAAGTGTAAGAAACAAGCAAGAAGACTGTGGTCTGGGGA 250
Db	30	GATGCGTCCCTTCAGTCAAGTGTAAGAAACAAGCAAGAAGACTGTGGTCTGGGGA 89
OY	251	GAGGTAAACCATTCCTTCCAACTGCTGCATGTCCCTGTGGTGAACAAGAACATGCC 310
Db	90	GAGGTAAACCATTCCTTCCAACTGCTGCATGTCCCTGTGGTGAACAAGAACATGCC 149
OY	311	TGCCCTCTGTCCACAGACAGACTGGGTAGTCCAAAGAAATCGGCAATGAGAGTGGCCCAG 370
Db	150	TGCCCTCTGTCCACAGACAGACTGGGTAGTCCAAAGAAATCGGCAATGAGAGTGGCCCAG 209
OY	371	GCGCTCTGTGTGGTTGCTCAACCTTGGACAACAAGCTAACACTGACGGGATTATCCT 430
Db	210	GCGCTCTGTGTGGTTGCTCAACCTTGGACAACAAGCTAACACTGACGGGATTATCCT 269
OY	431	TGAGAGAGAGAGAGATGCTGTGGGCTTGGAGACTACCAAAGGCTTGCTTTATAATTG 490
Db	270	TGAGAGAGAGAGAGATGCTGTGGGCTTGGAGACTACCAAAGGCTTGCTTTATAATTG 329
OY	491	TCTGTTTAGTTTTGGAAATTCCTACAAATTAGATAATTGTTAAAAATGGCCTTTCT 550
Db	330	TCTGTTTAGTTTTGGAAATTCCTACAAATTAGATAATTGTTAAAAATGGCCTTTCT 389
OY	551	AACCTCGGTGTGTGTGTGTGATAGCAATGACATAGAAAGCGGAACAACAGAAAATGATC 610
Db	390	AACCTCGGTGTGTGTGTGTGATAGCAATGACATAGAAAGCGGAACAACAGAAAATGATC 449
OY	611	TTGTGTTATCTGTACCACGACT--GGAAACAATTGTGTTCACAGAAACAATGTGTTG 667
Db	450	TTGTGTTATCTGTACCACGACTGTGGGAACATGTGTTCACAGAAACAATGTGTTG 509
OY	668	TTTATGCTTGAAGGTTAAAAATAGATAAGCAATGTAACATACTAATAATTAATSCAT 727
Db	510	TTTATGCTTGAAGGTTAAAAATAGATAAGCAATGTAACATACTAATAATTAATSCAT 569
OY	728	GAAAAGCCGACCTCCCTTAATCCTTTTGTGTGTGGAGAGAGAGCAAGCGAGGCCACCTG 787
Db	570	GAAAAGCCGACCTCCCTTAATCCTTTTGTGTGTGGAGAGAGAGCAAGCGAGGCCACCTG 629

QY 788 CTGCTTCATTTCTGCTGATGAGATTTTAACCTGCACTGATGAGAGCGTAACTGT 847  
 |||||  
 DB 630 CTCTCTTCAATTCCTGCTGATGAGATTTTAACCTGCACTGATGAGAGCGTAACTGT 689  
 |||||  
 QY 848 CCGGTAACCTGTAATATGCGTAACTGTCGCGGTAAACGGCTTTGCTCTGACTTCTCA 907  
 |||||  
 DB 690 CCGGTAACCTGTAATATGCGTAACTGTCGCGGTAAACGGCTTTGCTCTGACTTCTCA 749  
 |||||  
 QY 908 TCTTTGACTTGGCCAGGAA-GCCTGATTTGTTCAACCACTTACTT--CTAAGAACTGTT 964  
 |||||  
 DB 750 TCTTTGACTTGGCCAGGAAAGCGCGGATTTGTTCCACCACTTACTTCTAAGAAAGCGGTT 809  
 |||||  
 QY 965 TTTCTGTTTGGCCAGGATTTGTTATG 990  
 |||||  
 DB 810 TTCCGCTTTTGGCAAGGTGGAATG 835  
 |||||

RESULT 14  
 CB209388 938 bp mRNA linear EST 05-FEB-2003  
 LOCUS CB209388  
 DEFINITION IMAGE:30245249 5', mRNA sequence.  
 ACCESSION CB209388  
 VERSION CB209388.1 GI:28250951  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://img.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Dr. David Rowe and Dr. Mina  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: NDAM0319 row: 9 column: 18  
 High quality sequence stop: 712.  
 Location/Qualifiers  
 1..938  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30245249"  
 /lab\_host="PH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 164"  
 /note="Vector: PCMV-SPORT.1.cdab; site 1: EcoRV; site 2:  
 NotI; Non-normalized full-length enriched library from  
 pooled mouse embryonic limb, maxilla and mandible, day  
 10.5 and 11.5 (size selected for the 0.5-1 kb fragments)  
 Cloned directionally, priming method: Oligo-dT. CDNA  
 enrichment: >1k bp. Average insert size 1.8k bp. Priming  
 sequence: 5'-GACTAGTTTGAATCGGAGCGGCGCCCTG 3'. Tissue  
 contributed by: David Rowe. Library constructed by Resgen,  
 Invitrogen Corp."

BASE COUNT 204 a 153 c 211 g 202 t 168 others  
 ORIGIN  
 Query Match 64.9%; Score 739.6; DB 14; Length 938;  
 Best Local Similarity 99.1%; Pred. No. 66-148;  
 Matches 755; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 67 CCGGAGCGGAGGCTCCAGTCCGGAGCGGCAAGATGTTCTCTCAAGAGTGAACGC 126  
 |||||  
 DB 1 CCGGAGCGGAGGCTCCAGTCCGGAGCGGCAAGATGTTCTCTCAAGAGTGAACGC 60  
 |||||

QY 127 GGTAGCCATGATGGAGCTGGGAGCGTTGAGTGCATACCTGTCATCTGAGGGTCCAGGT 186  
 |||||  
 DB 61 GGTAGCCATGATGGAGCTGGGAGCGTTGAGTGCATACCTGTCATCTGAGGGTCCAGGT 120  
 |||||  
 QY 187 GATGAGTGCCTGCTTCCGATGTCGATGCTGAGTAAACAGAGAGAGCTGTGTGTGCTG 246  
 |||||  
 DB 121 GATGAGTGCCTGCTTCCGATGTCGATGCTGAGTAAACAGAGAGAGCTGTGTGTGCTG 180  
 |||||  
 QY 247 GGGAGAGTGAACCATCTCTTCCACAGCTGCTGCAATGCTGCTGCTGCTGCTGCTG 306  
 |||||  
 DB 181 GGGAGAGTGAACCATCTCTTCCACAGCTGCTGCAATGCTGCTGCTGCTGCTGCTG 240  
 |||||  
 QY 307 TCGCTGCCCTCTGTCGACAGAGAGTGGGTAGTCCAAAGAAATCGGCAATGAGAGTGGC 366  
 |||||  
 DB 241 TCGCTGCCCTCTGTCGACAGAGAGTGGGTAGTCCAAAGAAATCGGCAATGAGAGTGGC 300  
 |||||  
 QY 367 CCAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426  
 |||||  
 DB 301 CCAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 |||||  
 QY 427 TCTTGAAGAGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486  
 |||||  
 DB 361 TCTTGAAGAGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 |||||  
 QY 487 TTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546  
 |||||  
 DB 421 TTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 |||||  
 QY 547 TCTTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606  
 |||||  
 DB 481 TCTTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 |||||  
 QY 607 GATCTTTGTTATCTGTAATCCCACTGCT--GGAACATGCTGTTCAAGAAATCTGTT 663  
 |||||  
 DB 541 GATCTTTGTTATCTGTAATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 |||||  
 QY 664 TGTGTTATGCTGTAAGGTTAAATATGTAAGAAATGTAAGAAATGTAAGAAATGTAAG 723  
 |||||  
 DB 601 TGTGTTATGCTGTAAGGTTAAATATGTAAGAAATGTAAGAAATGTAAGAAATGTAAG 660  
 |||||  
 QY 724 CATTGAAAAGCCGACCTCTCTTATCTCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 783  
 |||||  
 DB 661 CATTGAAAAGCCGACCTCTCTTATCTCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 720  
 |||||  
 QY 784 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825  
 |||||  
 DB 721 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 762  
 |||||

RESULT 15  
 BY710626 965 bp mRNA linear EST 16-DEC-2002  
 LOCUS BY710626  
 DEFINITION BY710626 RIKEN full-length enriched, 10 days embryo Mus musculus  
 CDNA clone 261005N22 5', mRNA sequence.  
 ACCESSION BY710626  
 VERSION BY710626.1 GI:27121853  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://img.nci.nih.gov/  
 Tissue Procurement: Dr. David Rowe and Dr. Mina  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: NDAM0319 row: 9 column: 18  
 High quality sequence stop: 712.  
 Location/Qualifiers  
 1..965  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30245249"  
 /lab\_host="PH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 164"  
 /note="Vector: PCMV-SPORT.1.cdab; site 1: EcoRV; site 2:  
 NotI; Non-normalized full-length enriched library from  
 pooled mouse embryonic limb, maxilla and mandible, day  
 10.5 and 11.5 (size selected for the 0.5-1 kb fragments)  
 Cloned directionally, priming method: Oligo-dT. CDNA  
 enrichment: >1k bp. Average insert size 1.8k bp. Priming  
 sequence: 5'-GACTAGTTTGAATCGGAGCGGCGCCCTG 3'. Tissue  
 contributed by: David Rowe. Library constructed by Resgen,  
 Invitrogen Corp."

BASE COUNT 204 a 153 c 211 g 202 t 168 others  
 ORIGIN  
 Query Match 64.9%; Score 739.6; DB 14; Length 938;  
 Best Local Similarity 99.1%; Pred. No. 66-148;  
 Matches 755; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 67 CCGGAGCGGAGGCTCCAGTCCGGAGCGGCAAGATGTTCTCTCAAGAGTGAACGC 126  
 |||||  
 DB 1 CCGGAGCGGAGGCTCCAGTCCGGAGCGGCAAGATGTTCTCTCAAGAGTGAACGC 60  
 |||||

REFERENCE  
 AUTHORS  
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons







```

/db_xref="taxon:10090"
/clone="IMAGE:6510542"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/notes="Vector: PCMV-SPORE1.1.cdb; Site 1: EcoRV; Site 2:
NotI. Cloned unidirectionally. Primer: Oligo dT. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC library."
BASE COUNT      232 a      255 c      254 g      229 t
ORIGIN

```

```

Query Match      63.1%; Score 719.4; DB 13; Length 970;
Best Local Similarity 93.2%; Pred. No. 1.3e-143;
Matches 797; Conservative 0; Mismatches 51; Indels 7; Gaps 4;

21  CCGACGTGAGAGAGCGGCGAGAGAGACCTGCGCTCTTCTTTCGACATCCGAGAGCGAGGCT 80
27  CCGACGTGAGAGAGCGGCGAGAGAGACCTGCGCTCTTCTTTCGACATCCGAGAGCGAGGCT 86
81  CCAAGTCGGAGAGCGGCGAGAGAGAGTGTCTCTCTCAAGAGTGAACGGGATGATGTGA 140
87  CCAAGTCGGAGAGCGGCGAGAGAGAGTGTCTCTCTCAAGAGTGAACGGGATGATGTGA 146
141 GCTGGACGTTGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 200
147 GCTGGACGTTGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206
201 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 260
207 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266
261 ATTCTCTTCAACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 320
267 ATTCTCTTCAACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 326
321 GCCAGCAGAGCTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 380
327 GCCAGCAGAGCTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 386
381 TGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 440
387 TGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 446
441 AGGATGCTGTGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 500
447 AGGATGCTGTGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 506
501 TTTGGGAAATTTCTTCAAAATTAAGTAAATTTGTTAAATTAAGTAAATTTGTTAAATTA 560
507 TTTGGGAAATTTCTTCAAAATTAAGTAAATTTGTTAAATTAAGTAAATTTGTTAAATTA 566
561 TGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 620
567 TGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 626
621 TGTACCCACGACT---GGAACATTTGTTCACAGAAAGACATGTTGTTGTTAAATTA 677
627 TGTACCCACGACT---GGAACATTTGTTCACAGAAAGACATGTTGTTGTTAAATTA 686
678 AGGATGTTAAATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAAATTA 737
687 AGGATGTTAAATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAAATTA 746
738 CTCCTCTCAATCTCTTTTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796
747 CTCCTCTCAATCTCTTTTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 806
797 TTTGCTGTG---AATGAGATTTTAACTGTGACATCAAGTGAAGAGAGCT---AAGTGTGAGGTA 853
807 TTTGCTGTGCGACGAGAGATTTTAACTGTGACATCAAGTGAAGAGAGCT---AAGTGTGAGGTA 866
854 AACTGTAAATATGCGC 868

```

Db 867 AACTGTAAATATGCGC 881

RESULT 18  
B0571112  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

B0571112 728 bp mRNA linear EST 19-JUN-2002  
UI-M-FB0-bx2-0-08-0-UI.r1 NIH\_MMAP\_FBO Mus musculus cDNA clone  
IMAGE:5715295 5', mRNA sequence.  
B0571112  
B0571112.1 GI:21474429

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC <http://mgi.nci.nih.gov/>.  
1 (bases 1 to 728)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-f@mail.nih.gov](mailto:cgaps-f@mail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pyx-5.

FEATURES  
Location/Qualifiers  
1..728

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5715295"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH MGC FBO"  
/note="Organ: brain; Vector: pyx-Asc. Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pyx-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
, is TGAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT

187 a 151 c 197 g 193 t

Query Match 62.0%; Score 707.2; DB 13; Length 728;  
Best Local Similarity 99.2%; Pred. No. 5.2e-141;  
Matches 722; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

37 CGAGAACCTCGTCTCTTCTTTCGACATCCGAGAGCGAGGCTCCAAAGTGGAGAGGCGA 96  
1 CGAGAACCTCGTCTCTTCTTTCGACATCCGAGAGCGAGGCTCCAAAGTGGAGAGGCGA 60  
97 CAAGATGTTCTCTTCAAGAGTGAAGCGGCTTACCAATGAGACCTGGAGAGGCTGAGAG 156  
61 CAAGATGTTCTCTTCAAGAGTGAAGCGGCTTACCAATGAGACCTGGAGAGGCTGAGAG 120



[illegible]

AUTHORS	Komaba,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Shibata,K., Ariyama,O., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamanoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kasaihiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	Kawai,J., Shingawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arikawa,T., Hara,A., Fukunishi,Y., Komno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Salto,T., Okazaki,Y., Gotohori,T., Bono,H., Kanekura,T., Salto,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Pleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikolaic,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,Y., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., But,C., Fletcher,C., Fujita,M., Gariboldi,M., Gastlrich,S., Hill,D., Hochmann,M., Hume,D.A., Kamlay,M., Lee,N.H., Lyons,P., Machionni,L., Mashima,J., Mazzarello,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-Oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L., Wyshak-Sorits,A., Yoshida,K., Haegawa,Y., Kawai,H., Kohetsuki,S. and Hayashizaki,Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 722)
PUBMED	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arikawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayasutani,N., Hiramoto,K., Hiraoaka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,Y., Koijima,Y., Komno,H., Kouda,M., Koys,S., Kurihara,C., Matsumoto,T., Miyazaki,A., Nishi,K., Nemuru,K., Numazaki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Ozato,H., Salto,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suizaki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JULY-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN). Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suheiho-cho, Tsukuba-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-research@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.  cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGGACGAGACGCATCCAGACTCTTTTTCCTTTTTTN 3']. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went



prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers  
 1..721  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1810023A07"  
 /sex="male"  
 /tissue\_type="pancreas"  
 /dev\_stage="10 day old"  
 /clone\_lib="RIKEN full-length enriched, 10 day old male pancreas"

## BASE COUNT

190 a 147 c 200 g 184 t

## ORIGIN

Query Match 60.3%; Score 687.2; DB 14; Length 721;  
 Best Local Similarity 99.0%; Pred. No. 1e-136;

Matches 713; Conservative 0; Mismatches 3; Indels 4; Gaps 2;

19 GGGCCGCGGAGAGAGGCGGAGAACCTCGCTCTTCTTCCGCACTCCGAGCGCAG 78  
 2 GGGCGGCGGAGAGAGGCGGAGAACCTCGCTCTTCTTCCGCACTCCGAGCGCAG 61  
 79 CTCGAAGTCGGAGAGGAGCAAGATGTTCTCTCAAGAAATGGAGCGGATGCCATG 138  
 62 CTCGAAGTCGGAGAGGAGCAAGATGTTCTCTCAAGAAATGGAGCGGATGCCATG 120  
 139 GAGCTGGAGCGTTGAGTGCATCTGTGCCATCTGCAGGGTCCAGTATGATGATCCTG 198  
 121 GAGCTGGAGCGTTGAGTGCATCTGTGCCATCTGCAGGGTCCAGTATGATGATCCTG 180  
 199 CCTTCATGTCAAGCTGAAAAACAAGACAGCATGTGTGTGTGTGGGAGAGTGA 258  
 181 CCTTCATGTCAAGCTGAAAAACAAGACAGCATGTGTGTGTGTGGGAGAGTGA 240  
 259 CCATTCCTTCCCAACTGCGTCATGTCCTGTGGGAGAAACAAGAAATCGCGCTCT 318  
 241 CCATTCCTTCCCAACTGCGTCATGTCCTGTGGGAGAAACAAGAAATCGCGCTCT 300  
 319 GTGCCAGCAGACTGGGTAGTCCAAAGATCGCAATGAGAGTGGCCAGCGCTCT 378  
 301 GTGCCAGCAGACTGGGTAGTCCAAAGATCGCAATGAGAGTGGCCAGCGCTCT 360  
 379 GGTGTGTGTGTGAGTCCCTGAGCAAGAGCTAAACATGCGAGGATTCATCTTGAAGAG 438  
 361 GGTGTGTGTGTGAGTCCCTGAGCAAGAGCTAAACATGCGAGGATTCATCTTGAAGAG 420  
 439 AGAGATGCTGTGCGCTTGTGAGACTCAGCAAGAGGCTTCTTATTAATTTGCTGTTA 498  
 421 AGAGATGCTGTGCGCTTGTGAGACTCAGCAAGAGGCTTCTTATTAATTTGCTGTTA 480  
 499 GTTTGGGAAATTTCTACAAATTAATTTGTTAAAAATGCGCTTCTTACCTCTG 558  
 481 GTTTGGGAAATTTCTACAAATTAATTTGTTAAAAATGCGCTTCTTACCTCTG 540  
 559 TGTGTGTGTGTGATAGCATGATAGAGAGGAGAACACAGAAATGATCTTTGTTTA 618  
 541 TGTGTGTGTGTGATAGCATGATAGAGAGGAGAACACAGAAATGATCTTTGTTTA 600

QY 619 TCTGACCCAGCACT---GGAACATGTGTTTACAGAGAACATGTTGTTATGCT 675  
 DB 601 TCTGACCCAGCACTGTGGAACTGTTGTTTACAGAGAACATGTTGTTATGCT 660  
 QY 676 TGAGGTTAAAAAATAGATTAACGAATGTTCAGTACCAATTAATGATTAAGAAC 735  
 DB 661 TGAGGTTAAAAAATAGATTAACGAATGTTCAGTACCAATTAATGATTAAGAAC 720

## RESULT 21

LOCUS

B1647378

## DEFINITION

603279634F1\_NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5319872 5', mRNA sequence.

## ACCESSION

B1647378

## VERSION

B1647378.1 GI:15561614

## KEYWORDS

EST.

## SOURCE

Mus musculus

## ORGANISM

Mus musculus (house mouse)

## REFERENCE

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Plate: LLM11811 row: C column: 09  
 High quality sequence stop: 731.

## FEATURES

source

1..731

## Location/Qualifiers

/organism="Mus musculus"

## /mol\_type="mRNA"

/strain="129, C57BL/6J, FVB/N"

## /db\_xref="taxon:10090"

/clone="IMAGE:5319872"

## /tissue\_type="tumor, gross tissue"

/dev\_stage="10 months"

## /lab\_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPOrt6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

## BASE COUNT

192 a 157 c 200 g 182 t

## ORIGIN

Query Match 60.2%; Score 686.6; DB 12; Length 731;  
 Best Local Similarity 98.8%; Pred. No. 1.3e-136;

Matches 724; Conservative 0; Mismatches 4; Indels 5; Gaps 3;

QY 21 CCGAGCTGAGAGAGCGGAGGAACCTGCTCTTCTTCCGCACTCCGAGAGCGAGGCT 80  
 DB 1 CCGAGCTGAGAGAGCGGAGGAACCTGCTCTTCTTCCGCACTCCGAGAGCGAGGCT 60  
 QY 81 CCAAGTCGGAGAGCGACAAGATGTTCTCTCAAGATGGAACCGGTAGCCATGTGA 140  
 DB 61 CCAAGTCGGAGAGCGACAAGATGTTCTCTCAAGATGGAACCGGTAGCCATGTGA 120  
 QY 141 GGTGGAGCTTGAAGTGCATACCTGTGCATCTGAGAGGTCCAGGTGATGATGCTGCC 200  
 DB 121 GGTGGAGCTTGAAGTGCATACCTGTGCATCTGAGAGGTCCAGGTGATGATGCTGCC 180

OY	201	TTGCATGTCAGAGCTGAAAAACAAGACAAGAGACTGTGTGTGTCTGGGAGAAAGTGTAAAC	260
Db	181	TTGCATGTCAAGCTGTAAACACAGACAGAGACTGTGTGTGTCTGGGAGAGTGTAAAC	240
OY	261	ATTCCTTTCACAACTGTGTGCATGTCTCCTGTGGGTGAAAACAGAACATTCGTGCCCTCTGT	320
Db	241	ATTCCTTTCACAACTGTGTGCATGTCTCCTGTGGGTGAAAACAGAACATTCGTGCCCTCTGT	300
OY	321	GCCACGACGAGACTGGGTAGTCCAAAGATTCGGCAATGAGAGGTGGCCACGGCCCTCTGG	380
Db	301	GCCACGACGAGACTGGGTAGTCCAAAGATTCGGCAATGAGAGGTGGCCACGGCCCTCTGG	360
OY	381	TGTGGTGTCTACCCCTGGACAAAGACTAAACACTGCAGGGATTCATCCTTGAGAGAGAG	440
Db	361	TGTGGTGTCTACCCCTGGACAAAGACTAAACACTGCAGGGATTCATCCTTGAGAGAGAG	420
OY	441	AGGATGCTGTGTGCGCCTTTGAGACTCACCAAGGCTTCTTATTAATTTGTCTGTAGT	500
Db	421	AGGATGCTGTGTGCGCCTTTGAGACTCACCAAGGCTTCTTATTAATTTGTCTGTAGT	480
OY	501	TTTGGGAAATTCCTTACATTAATTAATTTGTTTAAAAATGGCCTTTTCTACCTCTGGTG	560
Db	481	TTTGGGAAATTCCTTACATTAATTAATTTGTTTAAAAATGGCCTTTTCTACCTCTGGTG	540
OY	561	TGTGTGTGTGATAGCAATGCAATGATAGAGAGCGAGAACACACAGAAAAATGATCTTGTGTTATC	620
Db	541	TGTGTGTGTGATAGCAATGCAATGATAGAGAGCGAGAACACACAGAAAAATGATCTTGTGTTATC	600
OY	621	TGTACCCACGACT--GGAACTTGTGTTCACAGAGAACATGTGTTTGTGTTATGCTTG	677
Db	601	TGTACCCACGACTGTGGGAACTGTGTTCACAGAGAACATGTGTTTGTGTTATGCTTG	659
OY	678	AGGTTTAAAAAATAGATTAACGAATGTTTACGTTAACAAATATAAATGCAATTTGAAAAAGCCGA	737
Db	660	AGGTTTAAAAAATAGATTAACGAATGTTTACGTTAACAAATATAAATGCAATTTGAAAAAGCCGA	718
OY	738	CTCCTCTTAATCC	750
Db	719	CTCCTCTTAATCC	731

RESULT 22  
 CBI96142  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source  
 CBI96142  
 AGENCECOURT 11258914 NIH MGC.135 Mus musculus cDNA clone  
 IMAGE:30137902 5', mRNA sequence.  
 CBI96142  
 CBI96142.1 GI:28223462  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 889)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Dr. David Rowe  
 cDNA library Preparation: Invitrogen Corp  
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM0039 row: n column: 23  
 High quality sequence stop: 718.  
 Location/Qualifiers  
 1..889  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"

[illegible]



```

/clone.lib="NCI CGAP Mam3"
/organism="mammal"; Vector: PCMV-Spore6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT      250 a      199 c      260 g      269 t
ORIGIN
Query Match      59.1%; Score 673.4; DB 12; Length 978;
Best Local Similarity 90.5%; Pred. No. 9.1e-134;
Matches 837; Conservative 0; Mismatches 61; Indels 27; Gaps 10;

QY 75 CAGGCTCCCAAGTCGGAGCGGAGCGAAGATGTTCTCTCTCAAGAGTGAACCGGATAGCA 134
Db 1 CAGGCTCCCAAGTCGGAGCGGAGCGAAGATGTTCTCTCTCAAGAGTGAACCGGATAGCA 60

QY 135 TGTGAGCTGGGACGTTGAGTGCGATACCTGTGCCATCTGCAGGGTCCAGGTGATGATG 194
Db 61 TGTGAGCTGGGACGTTGAGTGCGATACCTGTGCCATCTGCAGGGTCCAGGTGATGATG 120

QY 195 CTTGCTTGTGATGTCAAGTGAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
Db 121 CTTGCTTGTGATGTCAAGTGAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

QY 255 GTTACCATTCCTCCCAACTGCTGCATGTCCTGTGGGTGAACAGAACAAATGCTGTC 314
Db 181 GTTACCATTCCTCCCAACTGCTGCATGTCCTGTGGGTGAACAGAACAAATGCTGTC 240

QY 315 CTCTGTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 374
Db 241 CTCTGTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

QY 375 TCCTGAGTGTGCTGAGACCTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
Db 301 TCCTGAGTGTGCTGAGACCTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 435 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 494
Db 361 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 495 TTTAGTGTGGGAAATTTCTTACATTAAGATTAATTTGTTAAATAGGCTTCTTACCT 554
Db 421 TTTAGTGTGGGAAATTTCTTACATTAAGATTAATTTGTTAAATAGGCTTCTTACCT 480

QY 555 CTGATGTGTGTGTGATGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 614
Db 481 CTGATGTGTGTGTGATGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540

QY 615 TTTATCTGTAACCAAGAGCT--GGAACATTGTGTTCAAGAGAGAGAGAGAGAGAG 671
Db 541 TTTATCTGTAACCAAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

QY 672 TGCTGAGGGGTTAAAAATAGATTAAGAGATG--TTACAGTAACAAATTAATTCATTGAA 730
Db 601 TGCTGAGGGGTTAAAAATAGATTAAGAGATGTTTACAGTAACAAATTAATTCATTGAA 660

QY 731 AAGCCGACTCTCTCTTAATCTTTTGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
Db 661 AAGCCGACTCTCTCTTAATCTTTTGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 720

QY 786 TGCTGTCTTCAATTTGCT--GGAATGAGATTTTAACTGACACTGATGAGAGAGAG 839
Db 721 TGCTGTCTTCAATTTGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

QY 840 GTTACCTGCGGG--TAACTGTAATATGAGCGTAA--CTGTGCGGTAACCGGCTTTGTC 893
Db 781 GTTACCTGCGGGTAACCTGTAATATGAGCGTAACTGTGCGGTAACCGGCTTTGTC 840

QY 894 T-CTGACTTCTTCATCTTT--GACTTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 947
Db 841 TCCTGACTTCTTCATCTTTGAACTTTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

```

```

QY 948 AGTTCTAAGAACTGTTTCTGTTT 972
Db 901 AGTTCTAAGAAACCGGTTCTGTTT 925

RESULT 25
CB318729
LOCUS
DEFINITION
AGENCY: 12403690 NIH MGC 165 Mus musculus cDNA clone
IMAGE: 30285241 5', mRNA sequence.
CB318729
VERSION
CB318729.1 GI:28842964
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 787)
NIH-MGC http://mgi.nci.nih.gov/
NATIONAL INSTITUTE OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
Unpublished
CONTACT: Robert Strausberg, Ph.D.
Email: cga@b6-1@mail.nih.gov
Tissue Procurement: Dr. Leslie L. Heckert
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDCM50 row: j column: 02
High quality sequence stop: 581.
FEATURES
source
1..787
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30285241"
/risue_type="primary cultures of Sertoli cells"
/lab_host="DH10B (T1-phage-resistant)"
/clone.lib="NIH_MGC_165"
/note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI
(ggcatatggcc); Site 2: SfiI (ggcgctcgcc); 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGGCCGAGGCGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4 kb
(range 0.6-3.5 kb). 15/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH_MGC library."
BASE COUNT      209 a      161 c      214 g      203 t
ORIGIN
Query Match      58.9%; Score 671; DB 14; Length 787;
Best Local Similarity 95.8%; Pred. No. 3e-133;
Matches 723; Conservative 0; Mismatches 25; Indels 7; Gaps 3;

QY 68 GGGAGCGAGGCTCCAAAGTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 127
Db 4 GGGAGCGAGGCTCCAAAGTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 63

QY 128 GTAGCATGTGAGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 187
Db 64 GTAGCATGTGAGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123

QY 188 ATGAGATGCTGCTTCAATGTCAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 247
Db 124 ATGAGATGCTGCTTCAATGTCAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 183

QY 248 GAGAGTGAACATTTCTTCCAAATGCTGCAAGTCCCTGTGGTGAACAGAACAT 307

```

```

Db      184 GGAAGTGAACCAATCTTCCACAACTGCTGATGTCCTGTGGTGAAGCAAAAT 243
Qy      308 CGCTGCCCTCTGTGCCAGCAGACTGGTATGCCAAAGATCGCAAAATAGAGTGCC 367
Db      244 CGCTGCCCTCTGTGCCAGCAGACTGGTATGCCAAAGATCGCAAAATAGAGTGCC 303
Qy      368 CAGGCCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 427
Db      304 CAGGCCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 363
Qy      428 CTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
Db      364 CTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
Qy      488 TTGCTGCTTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 547
Db      424 TTGCTGCTTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 483
Qy      548 CCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 607
Db      484 CCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 543
Qy      608 ATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 664
Db      544 ATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 603
Qy      665 GTGTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 724
Db      604 GTGTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 663
Qy      725 ATTGAAGACCGACTCTCTCTTAATCTTTTGTG--TTGGAGAGAGCGAG--CGAGCG 780
Db      664 ATTGAAGACCGACTCTCTCTTAATCTTTTGTG--TTGGAGAGAGCGAG--CGAGCG 723
Qy      781 CACCTGCTGTCTTCAATTTGCTGTGAATGAGATT 815
Db      724 CACCCGCGCTGTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 758

```

## RESULT 26

```

Bg917948      916 bp      mRNA      linear      EST 05-JUN-2001
LOCUS      602820877F1 NCI_CGAP_Mamc Mus musculus cDNA clone IMAGE:4949851 5'
DEFINITION      mRNA sequence.
ACCESSION      Bg917948
VERSION      Bg917948.1 GI:14298424
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 916)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10903 row: i column: 20
High quality sequence stop: 760.
Location/Qualifiers
1..916

```

## FEATURES

```

source
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4949851"

```

```

/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mamc"
/notes="Organ: mammary; Vector: pCMV-Sport6; Site: 1; SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      232 a      202 c      264 g      218 t
ORIGIN

```

```

Query Match      58.4%; Score 665.4; DB 12; Length 916;
Best Local Similarity 93.8%; Pred. No. 4.7e-132;
Matches 761; Conservative 0; Mismatches 36; Indels 14; Gaps 6;

```

```

Qy      1 GTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db      40 GTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99
Qy      61 GCACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db      100 GCACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 159
Qy      121 GAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
Db      160 GAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 219
Qy      181 CAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db      220 CAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 279
Qy      241 GGTCTGGGAGAGGTACCAATCTCTCCACAACTGCTGCAATGCTGCTGCTGCTGCTG 300
Db      280 GGTCTGGGAGAGGTACCAATCTCTCCACAACTGCTGCAATGCTGCTGCTGCTGCTG 339
Qy      301 GAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db      340 GAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 399
Qy      361 GGTGCGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db      400 GGTGCGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459
Qy      421 GATTATCTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db      460 GATTATCTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
Qy      481 TATTATTTGTCTGTTAGTTTGGAAATCTTACAAATTAAGTAAATTTGTTAAAT 540
Db      520 TATTATTTGTCTGTTAGTTTGGAAATCTTACAAATTAAGTAAATTTGTTAAAT 579
Qy      541 GGCCTTTCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
Db      580 GGCCTTTCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 638
Qy      601 GAAATGATCTTTGTTTATCTGTACCAAGAT--GGAACCTGTGTCTCAAGAAAC 657
Db      639 GAAATGATCTTTGTTTATCTGTACCAAGAT--GGAACCTGTGTCTCAAGAAAC 698
Qy      658 ATTGTTTGTGTTAGCTTGAAGGTT--AAAAATGATTAAGAGTGTACAGTAAACAA 716
Db      699 CATG--TGAGGTTATGCTTGAAGGTTAAATAATGATTAAGAGTGTACAGTAAACAA 756
Qy      717 TAAAA-----TGATGAAAAAGCGACTCTCTTAATCTTTTGTGT--TGAGAGAG 769
Db      757 ATAAAAATGCGGTGAAAAAGCGCAATCTCTTAATCGTTGTAGTGAAGGAGAGAG 816
Qy      770 GGAACGAGGCAACCTGCTGCTTCAATTTG 800
Db      817 GGAACGAGGCAACCTGCTGCTTCAATTTG 847

```



```

RESULT 27
CA977674          944 bp  mRNA  linear  EST 06-JAN-2003
LOCUS             CA977674
DEFINITION       AGENCOURT_11293247 NIH_MGC_164 Mus musculus cDNA clone
IMAGE:30146934 5', mRNA sequence.
ACCESSION        CA977674
VERSION          CA977674.1 GI:27510328
KEYWORDS         EST.
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 944)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cgabbs-r@mail.nih.gov
                  Tissue Procurement: Dr. David Rowe and Dr. Mina
                  CDNA Library Preparation: Invitrogen Corp
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution information can be
                  found through the I.M.A.G.E. Consortium/ILNL at:
                  http://image.llnl.gov
                  Plate: NDAM0063 row: 9 column: 07
                  High quality sequence stop: 673.
FEATURES
    source
        1..944
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:30146934"
            /lab_host="DH10B (phage-resistant)"
            /clone_1ib="NIH MGC 164"
            /note="Vector: PCMV-SpOBT.1.ccdB; Site 1: EcoRV; Site 2:
            NOTI; Non-normalized full-length enriched library from
            pooled mouse embryonic limb, maxilla and mandible, day
            10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
            Cloned directionally, priming method: Oligo-dT. CDNA
            enrichment: >1k bp. Average insert size 1.8k bp. Priming
            sequence: 5'-GACTAGTCTGATCGGAGCGGCCCT(T) 3'. Tissue
            contributed by, David Rowe. Library constructed by Resgen,
            Invitrogen Corp."
BASE COUNT      216 a 162 c 242 g 208 t 116 others
ORIGIN
Query Match      58 3%; Score 664.6; DB 14; Length 944;
Best Local Similarity 86.8%; Pred. No. 7e-132;
Matches 683; Conservative 0; Mismatches 101; Indels 3; Gaps 1;
QY      21  CCGACGTGAGAGACGGCGAGGAACCTGCGCTCTTCTTGCACTCCGGAGCGCAGGCT 80
Db      1  CCGACGTGAGAGACGGCGAGGAACCTGCGCTCTTCTTGCACTCCGGAGCGCAGGCT 60
QY      81  CCAAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGGAGCGGCTGACATGGA 140
Db      61  CCAAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGGAGCGGCTGACATGGA 120
QY      141 GCTGGGAGCTTGAATGGATCTGCTGTCATCTGCGAGGTCCAGTGATGATGCTGCC 200
Db      121 GCTGGGAGCTTGAATGGATCTGCTGTCATCTGCGAGGTCCAGTGATGATGCTGCC 180
QY      201 TTGCATGTCAAGCTGAAAAACAAGACAAGAGACTGTGTTGTGCTGGGAGAGTGAAC 260
Db      181 TTGCATGTCAAGCTGAAAAACAAGACAAGAGACTGTGTTGTGCTGGGAGAGTGAAC 240
QY      261 ATTCTTCCCAACTTGTCTGCTATCCCTGTGGGGTGAACAACAATCGCTGCCCTGT 320
Db      241 ATTCTTCCCAACTTGTCTGCTATCCCTGTGGGGTGAACAACAATCGCTGCCCTGT 300
QY      321 GCCAGCAGCACTGGGTAGTCCAAAGAAATCGCAAAATGAGAGGTGCCAGGCGCTCTGG 380

```

```

Db      301 GCCAGCAGCACTGGGTAGTCCAAAGAAATCGCAAAATGAGAGGTGCCAGGCGCTCTGG 360
QY      381 TGTGTTGTGACCTCGGACAAAGACTAAACACTGCGAGGGGATTCATCTTGAAGAGAG 440
Db      361 TGTGTTGTGACCTCGGACAAAGACTAAACACTGCGAGGGGATTCATCTTGAAGAGAG 420
QY      441 AGGATGTGTGGCCCTTTGAGACTCAACAAAGCTGCTTATTAATTTGTCTGTAGT 500
Db      421 AGGATGTGTGGCCCTTTGAGACTCAACAAAGCTGCTTATTAATTTGTCTGTAGT 480
QY      501 TTGGGAAATTCCTCAATTAAGATAATTTGTTAAATGAGGCTTCTTACCTGTG 560
Db      481 TTGGGAAATTCCTCAATTAAGATAATTTGTTAAATGAGGCTTCTTACCTGTG 540
QY      561 TGTGTGTGATACGAATGATGAAGAGGAGAAACACCGAAATGATCTTGTATC 620
Db      541 TGTGTGTGATACGAATGATGAAGAGGAGAAACACCGAAATGATCTTGTATC 600
QY      621 TGTACCCAGACT---GGAATTTGTGTGACAGAGAACATTTGTTATGCTTG 677
Db      601 TGTACCCAGACTGTGGGAAATTTGTTTACAGAGAACATTTGTTATGCTTG 660
QY      678 AGGTTAAAAAATGATTAACGATGTTACAGTAACAAATAAATGATTAAGGCGGA 737
Db      661 AGGTTAAAAAATGATTAACGATGTTACAGTAACAAATAAATGATTAAGGCGGA 720
QY      738 CTCTCTCTAATCTTTTGTGTGGGAGAGAGCAAGCAGGACCTCTGCTTCAT 797
Db      721 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 780
QY      798 TTGCGGT 804
Db      781 TTTTAT 787

```

```

RESULT 28
BO177475          676 bp  mRNA  linear  EST 30-APR-2002
LOCUS             BO177475
DEFINITION       UT-M-ER0-bwo-n-12-0-UT-1 NIH BMP2_ER0 Mus musculus cDNA clone
IMAGE:5699915 5', mRNA sequence.
ACCESSION        BO177475
VERSION          BO177475.1 GI:20352967
KEYWORDS         EST.
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 676)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cgabbs-r@mail.nih.gov
                  Tissue Procurement: Dr. James Lin, University of Iowa
                  CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
                  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                  Clone distribution information can be
                  found through the I.M.A.G.E. Consortium/ILNL at:
                  http://image.llnl.gov
                  This clone was contributed by the Brain Molecular Anatomy Project
                  (BMAP)
FEATURES
    source
        1..676
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="CS7BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:5699915"
            /tissue_type="whole brain"
            /dev_stage="embryo 15.5 dpc"
            /lab_host="DH10B (T1 phage resistant)"

```



/clone lib="NIH\_BMAP\_ER0"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then  
 cloned directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GTGGGTGGA. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP): 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institutes of Mental Health  
 (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 167 a 143 c 192 g 174 t  
 ORIGIN  
 Query Match 57.3%; Score 653.4; DB 13; Length 676;  
 Best Local Similarity 99.4%; Pred. No. 1.7e-129;  
 Matches 667; Conservative 0; Mismatches 1; Indels 3; Gaps 1;  
 QY 21 CCGACGTGAGGACGCGCGAGGAACTCTGCTCTTCTTCCGACCTCCGGAGCGCAGGCT 80  
 Db 1 CCGACGTGAGGACGCGCGAGGAACTCTGCTCTTCTTCCGACCTCCGGAGCGCAGGCT 60  
 QY 81 CCAAGTCGGGAGCGGCAAGATGTTCTCTCAAGAGTGGAACGGGGTACCATGTGA 140  
 Db 61 CCAAGTCGGGAGCGGCAAGATGTTCTCTCAAGAGTGGAACGGGGTACCATGTGA 120  
 QY 141 GCTGGGACGTTGAGTGCATCTGTCATCTGCGAGGCGGCGGATGATGCTGTC 200  
 Db 121 GCTGGGACGTTGAGTGCATCTGTCATCTGCGAGGCGGCGGATGATGCTGTC 180  
 QY 201 TTTCGATGTCAGGTGAAAACAAGAGAGAGCTGTGTGTGCTGGGAGAGTGTAC 260  
 Db 181 TTTCGATGTCAGGTGAAAACAAGAGAGAGCTGTGTGTGCTGGGAGAGTGTAC 240  
 QY 261 ATTCTCTCCACAATCTGCTGATGCTCTGTGGGTGAAACAAGCAATCCGCTCTGT 320  
 Db 241 ATTCTCTCCACAATCTGCTGATGCTCTGTGGGTGAAACAAGCAATCCGCTCTGT 300  
 QY 321 GCCAGAGAGTGGGTGATCCAAAGATGGGCAATGAGAGGGGCCAGGGGCTGCTGG 380  
 Db 301 GCCAGAGAGTGGGTGATCCAAAGATGGGCAATGAGAGGGGCCAGGGGCTGCTGG 360  
 QY 381 TGTGTGTGTCGACCTTGACAAAGACTAACTGCGAGGGATTCATCTTGAAGAGAG 440  
 Db 361 TGTGTGTGTCGACCTTGACAAAGACTAACTGCGAGGGATTCATCTTGAAGAGAG 420  
 QY 441 AGGATGCTGGGCGCTTGAAGTCAACAAGGCTGTTATTAATTTGTCTGTTAGT 500  
 Db 421 AGGATGCTGGGCGCTTGAAGTCAACAAGGCTGTTATTAATTTGTCTGTTAGT 480  
 QY 501 TTGGGAAATTCCTCAATTAAGATTAATTTGTTAAATGGCTTCTTACTCTGGTG 560  
 Db 481 TTGGGAAATTCCTCAATTAAGATTAATTTGTTAAATGGCTTCTTACTCTGGTG 540  
 QY 561 TGTGTGTGTCATGATGATGAGAGAGAGACCAAGAAATGATCTTTGTTATC 620  
 Db 541 TGTGTGTGTCATGATGATGAGAGAGAGACCAAGAAATGATCTTTGTTATC 600  
 QY 621 TGTACCAAGACT--GAAACATTTGTTTCAAGAAACATTTGTTGTTATGCTTG 677  
 Db 601 TGTACCAAGACTGTTGGAAACATTTGTTTCAAGAAACATTTGTTGTTATGCTTG 660  
 QY 678 AGGGTTAAAA 688  
 Db 661 AGGGTTAAAA 671

RESULT 29

BQ044727  
 LOCUS BQ044727 686 bp mRNA linear EST 28-MAR-2002  
 DEFINITION U1-M-EMO-bum-j-14-0-UI.r1 NIH\_BMAP\_EM0 Mus musculus cDNA clone  
 IMAGE:5694061 5', mRNA sequence.  
 ACCESSION BQ044727  
 VERSION BQ044727.1 GI:19795492  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Euteleostomi; Rodentia; Scleroglossa; Muridae; Mus.  
 REFERENCES  
 1 (bases 1 to 686)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 COMMENT  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/HLN at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pYX-5.  
 Location/Qualifiers  
 1..686  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5694061"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 18.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_EM0"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then  
 cloned directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GAGCAGGAC. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP): 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institute of Mental Health  
 (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 179 a 135 c 174 g 197 t  
 ORIGIN  
 Query Match 57.2%; Score 651.6; DB 12; Length 686;  
 Best Local Similarity 98.7%; Pred. No. 4.2e-129;  
 Matches 678; Conservative 0; Mismatches 5; Indels 4; Gaps 2;  
 QY 231 ACTGTGTGTGTCGGGAGAGTGAACCATCTTCCACAAGCTGCTGCTGCT 290  
 Db 1 ACTGTGTGTGTCGGGAGAGTGAACCATCTTCCACAAGCTGCTGCTGCTGCT 60  
 QY 291 GGGTGAACAGAAACATGCTGCTCTGTGTCAGAGAGCTGGTGTGCAAGATG 350  
 Db 61 GGGTGAACAGAAACATGCTGCTCTGTGTCAGAGAGCTGGTGTGCAAGATG 120  
 QY 351 GCAATGAGAGGTGGCGCCAGAGGCTCTGTGTGTGTGTCACCTTGAACAAGACTAA 410  
 Db 121 GCAATGAGAGGTGGCGCCAGAGGCTCTGTGTGTGTGTCACCTTGAACAAGACTAA 180  
 QY 411 CACTGACGAGGATCATCTTGAGAGAGAGAGATGTGTGTGCTTGTGACTACCA 470







QY 452 CGCCTTGAGACTCACCAGGCTGCTTATTAATTTGCTGTTAGTTTGGGAATT 511  
 Db 362 CGCCTTGAGACTCACCAGGCTGCTTATTAATTTGCTGTTAGTTTGGGAATT 421  
 QY 512 CTCCTCAATTAAGATTAATTTGTTAAAAATGCGCTTCTCACTGCTGCTGTGTGA 571  
 Db 422 CTCCTCAATTAAGATTAATTTGTTAAAAATGCGCTTCTCACTGCTGCTGTGTGA 481  
 QY 572 TACGATGATAGAGAGGAGAACCCAGAAATATGATCTTTGTTATCTGTAACCA 631  
 Db 482 TACGATGATAGAGAGGAGAACCCAGAAATATGATCTTTGTTATCTGTAACCA 541  
 QY 632 CT---GGAATTTGTTGACAGAAATGTTGTTGTTATGCTGAGGGTTAAAA 688  
 Db 542 CTGTGGAAACATTTGTTGACAGAAATGTTGTTGTTATGCTGAGGGTTAAAA 601  
 QY 689 ATAGATTAACGATGTTAGATTAACAAATTAATGATTAAGAGCGAGCTCTCTAAT 748  
 Db 602 ATAGATTAACGATGTTAGATTAACAAATTAATGATTAAGAGCGAGCTCTCTAAT 661  
 QY 749 CCTTTT 755  
 Db 662 CCTTTT 668

RESULT 34  
 B1556918 737 bp mRNA linear EST 05-SEP-2001  
 LOCUS 60339970F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5292832 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1556918  
 VERSION B1556918.1 GI:15444232  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 737)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L14M11740 row: 1 column: 17  
 High quality sequence scop: 685.  
 Location/Qualifiers  
 1. 737

## FEATURES

source

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="129, C57BL/6J, FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5292832"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="10 months"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCI\_CGAP\_Mam3"  
 /note="Organ: mammary; Vector: pCMV-SPORT; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH  
 Reference for transgenic model: Xu et al., Nature Genetics  
 22, 37-43 (1999)."  
 BASE COUNT 200 a 152 c 212 g 173 t  
 ORIGIN

Query Match 56.7%; Score 646.2; DB 12; Length 737;  
 Best Local Similarity 97.1%; Pred. No. 6.1e-128;  
 Matches 702; Conservative 0; Mismatches 13; Indels 8; Gaps 4;

QY 20 GCCGACGTGAGAGACGGAGAGAACCTGCGTCTTTCTTCGACACTCCGGAGCGAGCG 79  
 Db 1 GCCGACGTGAGAGACGGAGAGAACCTGCGTCTTTCTTCGACACTCCGGAGCGAGCG 60  
 QY 80 TTCAGTGGGAGGCGACAAATGTTCTCTCAAGAAAGTGAACCGGTACCATGTGG 139  
 Db 61 TCCAAATGGGGAGGCGACAAAGATGTTCTCTCAAGAAAGTGAACCGGTACCATGTGG 120  
 QY 140 AGCTGGAGCGTGAATGCGATCCTGTGCAATCTGCAAGGTCAGGTATGGAATGCGTCG 199  
 Db 121 AGCTGGAGCGTGAATGCGATCCTGTGCAATCTGCAAGGTCAGGTATGGAATGCGTCG 180  
 QY 200 CTTGATGTCAGCTGAAACCAAGCAAGAGACTGTGTGTGCTGTGGAGAGTGAAC 259  
 Db 181 CTTGATGTCAGCTGAAACCAAGCAAGAGACTGTGTGTGCTGTGGAGAGTGAAC 240  
 QY 260 CATTCCTTCACAACTGCTGCATGTCCTGTGTGTGGAACAGAAATGCTGCTGCTG 319  
 Db 241 CATTCCTTCACAACTGCTGCATGTCCTGTGTGTGGAACAGAAATGCTGCTGCTG 300  
 QY 320 TGCCAGACGAACTGGGTAGTCCAAAGAAATGCGCAATGAGAGTGGCCGAGGCTGCTG 379  
 Db 301 TGCCAGACGAACTGGGTAGTCCAAAGAAATGCGCAATGAGAGTGGCCGAGGCTGCTG 360  
 QY 380 GTGTGTTGCTGACCTCGGACCAAGACTTAACTGCAAGGAGTTCATCTTTGAGAGAGA 439  
 Db 361 GTGTGTTGCTGACCTCGGACCAAGACTTAACTGCAAGGAGTTCATCTTTGAGAGAGA 420  
 QY 440 GAGATGCTGTGTGCGCTTTGAGACTCACCAGAAAGCTGCTTATTAATTTGCTGTTAG 499  
 Db 421 GAGATGCTGTGTGCGCTTTGAGACTCACCAGAAAGCTGCTTATTAATTTGCTGTTAG 480  
 QY 500 TTTTGGAAATCTCTACAAATTAATTAATTTGTT-AAAAATGGCTTCTTCACTCTGCG 558  
 Db 481 TTTTGGAAATCTCTTCAATTAATTAATTTGTTAAATTAATTAATTTGTTCTTCACTCTGCG 540  
 QY 559 TGTGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618  
 Db 541 TGTGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 QY 619 TGTGTATCCACGACT---GGAACATTGTTGTTACAGAAAGACATTGTTGTTATGCT 675  
 Db 601 TGTGTATCCACGACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 657  
 QY 676 TGAGGTTTAAAAATGATTAACGAAATGTTACAGTAACAAATTAATTAATTAATTAATTAAT 734  
 Db 658 TGAGGTTTAAAAATGATTAACGAAATGTTACAGTAACAAATTAATTAATTAATTAATTAAT 717  
 QY 735 CGA 737  
 Db 718 CGA 720

RESULT 35  
 B1737338 642 bp mRNA linear EST 17-DEC-2002  
 LOCUS B1737338 RIKEN full-length enriched, 14 days pregnant adult female  
 DEFINITION placenta Mus musculus cDNA clone IS30025D07 5', mRNA sequence.  
 ACCESSION B1737338  
 VERSION B1737338.1 GI:27150465  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 642)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,

Gojobori, T., Baldarelli, R., Hill, D. P., Bulf, C., Hume, D. A., Quackenbush, U., Schriml, L. M., Kanapin, A., Matunda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradd, D., Brusic, V., Chochic, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, F., Pavan, W. J., Pereira, G., Pesole, G., Petkovsky, N., Pillel, R., Pontius, J. U., Qi, D., Ring, Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setbon, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wymshaw-Borrie, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komoto, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Alizawa, K., Arikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL MEDLINE  
PUBMED  
22354683  
12466851

COMMENT  
Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Horii, F., Imotani, K., Iishi, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Komoto, H., Koyama, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

Location/Qualifiers

1. 642

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1530025D07"

/sex="female"

/tissue\_type="placenta"

/dev\_stage="14 days pregnant adult"

BASE COUNT 151 a 145 c 187 g 158 t 1 others  
ORIGIN /clone lib="RIKEN full-length enriched, 14 days pregnant adult female placenta"

Query Match 55.2%; Score 629; DB 14; Length 642;  
Best Local Similarity 99.8%; Pred. No. 2,9e-124;  
Matches 629; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

5 TGGCGCCGCCCATGCGCCGACGTGAGAGACGGCGAGAACTCTGCTCTTCTTCGAC 64

1 TGGCGCCGCCCATGCGCCGACGTGAGAGACGGCGAGAACTCTGCTCTTCTTCGAC 60

65 TCCGGAGGCGGACGCTCCCAAGTCGGAGGCGCAAGATTTCTCTCAAGAAGTGAAC 124

61 TCCGGAGGCGGACGCTCCCAAGTCGGAGGCGCAAGATTTCTCTCAAGAAGTGAAC 120

125 GCGGTAGCATGTGAGAGTGGAGCTTGAAGTCACTGTCATCTGCAAGGTCAG 184

121 GCGGTAGCATGTGAGAGTGGAGCTTGAAGTCACTGTCATCTGCAAGGTCAG 180

185 GTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244

181 GTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240

245 TGGGGAGAGTGAACCATTCCTCCCAACTGCTGATGCTGCTGCTGCTGCTGCTGCT 304

241 TGGGGAGAGTGAACCATTCCTCCCAACTGCTGATGCTGCTGCTGCTGCTGCTGCT 300

305 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 364

301 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360

365 GCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 424

361 GCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420

425 CATCTTGAAG 484

421 CATCTTGAAG 480

485 AATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544

481 AATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

545 TTCTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 604

541 TTCTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

605 ATGATCTTGTATATCTGACCCAGACTG 634

601 ATGATCTTGTATATCTGACCCAGACTG 630

RESULT 36  
CB209237 920 bp mRNA linear EST 05-FEB-2003

LOCUS AGENCOURT 11340479 NIH-MGC 164 Mus musculus cDNA clone

DEFINITION IMAGE:30243509 5', mRNA sequence.

CB209237

VERSION CB209237.1 GI:28250800

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC http://mgi.nci.nih.gov/

Unpublished National Institutes of Health, Mammalian Gene Collection (MGC)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue Procurement: Dr. David Rowe and Dr. Mina

CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM0314 row: 0 column: 06  
 High quality sequence stop: 644.

## FEATURES

source

1..920  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30243509"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_164"  
 /note="Vector: pCMV-Sport6.1.cdab; Site 1: EcoRV; Site 2:  
 NotI; Non-normalized full-length enriched library from  
 pooled mouse embryonic limb, maxilla and mandible, day  
 10.5 and 11.5 (size selected for the 0.5-1 kb fragments)  
 Cloned directionally, priming method: Oligo-dT. CDNA  
 enrichment: >1k bp. Average insert size 1.8k bp. Priming  
 sequence: 5'-GACTAGTTTATGATGCGAGCGGCCCTTT-3'. Tissue  
 contributed by, David Rowe. Library constructed by Resgen,  
 Invitrogen Corp."

BASE COUNT 184 a 152 c 200 g 194 t 190 others

## ORIGIN

Query Match 55.0%; Score 627.2; DB 14; Length 920;  
 Best Local Similarity 98.9%; Pred. No. 7.1e-124;  
 Matches 642; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

48 GGGCTCTTCTTCCGCACTCCGGAGGCGGAGGCGGCAAGTCCGAGGCGCAAGATTTCT 107  
 1 GGGCTCTTCTTCCGCACTCCGGAGGCGGAGGCGGCAAGTCCGAGGCGCAAGATTTCT 60  
 108 CTCTCAAGAAGTGAACGCGGTAGCATGTGAGCTGGGACGTTGAGTGGATCTGTG 167  
 61 CTCTCAAGAAGTGAACGCGGTAGCATGTGAGCTGGGACGTTGAGTGGATCTGTG 120  
 168 CCATCTGCAAGGCTCAGGTGATGATGCTGCTTGCATGTCAGTGAAGTGAAGCAAG 227  
 121 CCATCTGCAAGGCTCAGGTGATGATGCTGCTTGCATGTCAGTGAAGTGAAGCAAG 180  
 228 AGGACGTGTGTGTGCTGGGAGAGTGAACCATCTTCCCACTGCTGATGTCCTCC 287  
 181 AGGACGTGTGTGTGCTGGGAGAGTGAACCATCTTCCCACTGCTGATGTCCTCC 240  
 288 TGTGGGTGAAGCAAGCAATCGCTGCTGTGCGCAGCAGCACTGGTATGTCAGAA 347  
 241 TGTGGGTGAAGCAAGCAATCGCTGCTGTGCGCAGCAGCACTGGTATGTCAGAA 300  
 348 TCGGCAATGAGAGTGGGCCGAGGCGCTCTGTGTGTGCTGACCTGAGCAAGACT 407  
 301 TCGGCAATGAGAGTGGGCCGAGGCGCTCTGTGTGTGCTGACCTGAGCAAGACT 360  
 408 AAAACATGAGAGGAGTTCATCTTGAAGAGAGAGATGTCGCCCTTTGAGACTCAG 467  
 361 AAAACATGAGAGGAGTTCATCTTGAAGAGAGAGATGTCGCCCTTTGAGACTCAG 420  
 468 CAAGGCTGCTTATTAATTTGTCTGTTAGTTTGGGAAATTTCTCAATTAAGATA 527  
 421 CAAGGCTGCTTATTAATTTGTCTGTTAGTTTGGGAAATTTCTCAATTAAGATA 480  
 528 ATTGTTAAATATGCTTTCTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 587  
 481 ATTGTTAAATATGCTTTCTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540  
 588 AGGAGAACACGAGAAATGATTTGTTTATTTGTTATGTTATGTTATGTTATGTTAT 644  
 541 AGGAGAACACGAGAAATGATTTGTTTATTTGTTATGTTATGTTATGTTATGTTAT 600  
 645 TTCACAGAAACATTTGTTGTTTATGTTATGTTATGTTATGTTATGTTATGTTAT 693

Db

601 TTCACAGAAACATTTGTTGTTTATGTTATGTTATGTTATGTTATGTTATGTTAT 649

RESULT 37  
 BF302267  
 LOCUS  
 DEFINITION  
 mRNA sequence.  
 accession  
 BF302267  
 version  
 BF302267.1 GI:11248790  
 keywords  
 EST.  
 source  
 Mus musculus (house mouse)  
 organism  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 1 (bases 1 to 822)  
 TITLE  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LAM9455 row: a column: 09  
 High quality sequence stop: 738.

## FEATURES

source

1..822  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4166288"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1lb="NCI\_CGAP\_SG2"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site 1:  
 NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
 dT. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 227 a 178 c 229 g 188 t

## ORIGIN

Query Match 54.8%; Score 624.4; DB 10; Length 822;  
 Best Local Similarity 96.9%; Pred. No. 2.8e-123;  
 Matches 713; Conservative 0; Mismatches 11; Indels 12; Gaps 7;

6 GCGCGCGCCGATGCGCGAGAGAGAGCGGAGGAACCTGCTCTTCTTCTGCACT 65  
 5 GCGTGGCGCCATGAGCCGACGAGAGAGCGGAGGAACCTGCTCTTCTTCTGCACT 64  
 66 CCGGAGCGGAGGCTCCAGTGGGAGGAGGAAGATGTTCTCTCAAGATGGAAG 125  
 65 CCGGAGCGGAGGCTCCAGTGGGAGGAGGAAGATGTTCTCTCAAGATGGAAG 124  
 126 CCGTACCATGTGAGAGCTGGAGAGTTGAGTGCATACCTGTCATCTGACAGGTCAG 185  
 125 CCGTACCATGTGAGAGCTGGAGAGTTGAGTGCATACCTGTCATCTGACAGGTCAG 184  
 186 TGATGATGCTGCTGCTTGCATGTCAAGCTGAAACAGACAGAGAGCTGTGTGTCT 245  
 185 TGATGATGCTGCTGCTTGCATGTCAAGCTGAAACAGACAGAGAGCTGTGTGTCT 244  
 246 GGGAGAGGTATACCATCTTCCCACTGTCATGTCATGTCCTGTGGGAGGAAGCA 305  
 245 GGGAGAGGTATACCATCTTCCCACTGTCATGTCATGTCCTGTGGGAGGAAGCA 304  
 306 ATCGCTGCTCTGTGCGCAGCAGAGCTGGGTATGTCGAAGATTCGCAATGAGAGTGG 365



RESULT 38					
BF579825					
LOCUS	6F02095651P1	NCI_CGAP_Co24	Mus musculus cDNA clone IMAGE:4215924	5'	'
DEFINITION					
ACCESSION	BF579825		755 bp	mRNA	linear
VERSION	BF579825				
KEYWORDS	BF579825.1	GI:11653537			
SOURCE	EST.				
ORGANISM	Mus musculus (house mouse)				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus. 1 (bases 1 to 755) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. <a href="mailto:rstra@nsl.nhl.gov">rstra@nsl.nhl.gov</a>				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					

## FEATURES

BASE COUNT

RESULT 39

ACCESSION	BE628111
VERSION	BE628111.1
	GI:9908559

REFERENCE	ORGANISM	SOURCE	FAVORABLE
1 (bases 1 to 620)	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.	Mus musculus (house mouse)	
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		



AUTHORS		NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
JOURNAL		Tumor Gene Index	
COMMENT		Unpublished	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgapbs-r@mail.nih.gov	
		This clone is available royalty-free through LNL; contact the	
		IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information.	
		MG1:1062944	
		High quality sequence stop: 456.	
FEATURES		Location/Qualifiers	
source		1..620	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:3373340"	
		/sex="female (lactating)"	
		/tissue_type="mammary gland"	
		/lab_host="MDH10B"	
		/clone_lib="Soares_mammary_gland_NMLMG"	
		/note="vector: pT7T3-Pac (pharmacia) with a modified	
		polylinker; 1st strand cDNA was prepared from mammary	
		gland and tissue from a lactating female, and was then primed	
		with a Not I - oligo(dT) primer. Double-stranded cDNA was	
		ligated to Eco RI adaptors (Pharmacia), digested with Not	
		I and cloned into the Not I and Eco RI sites of the	
		modified pT7T3 vector. Library is normalized. Library	
		was constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		163 a 167 c 118 g 171 t 1 others	
ORIGIN			
Query Match		52.4%; Score 597.2, DB 10; Length 620;	
Best Local Similarity		98.9%; Pred. No. 1.9e-117;	
Matches 612; Conservative		0; Mismatches 4; Indels 3; Gaps 1;	
QY	122 AACGGGATGACCATGTGGAGCGTGGATGATCGATACCTGCGCATGCGAGGGCTC	181	
DB	620 AACGGGATGACCATGTGGAGCGTGGATGATCGATACCTGCGCATGCGAGGGCTC	561	
QY	182 CAGGTGATGATGCTGCTGCTTCGATGTCAGCTGAAAACAAGACAGAGACTGTGTGTG	241	
DB	560 CAGGTGATGATGCTGCTGCTTCGATGTCAGCTGAAAACAAGACAGAGACTGTGTGTG	501	
QY	242 GTCTGGGAGAGCTGTAACCATTCCTTCCAAACCTGCTGATGCTCCCTGTGGTGAACAG	301	
DB	500 GTCTGGGAGAGCTGTAACCATTCCTTCCAAACCTGCTGATGCTCCCTGTGGTGAACAG	441	
QY	302 AACATGCTGCTGCTGCTGCTGCGACAGAGACCTGGATGTCGAAAGATGGCAATGAAGAG	361	
DB	440 AACATGCTGCTGCTGCTGCTGCGACAGAGACCTGGATGTCGAAAGATGGCAATGAAGAG	381	
QY	362 GTGGCCACAGCGCTCTGTGTGTGTGTGATCGAACCCTGAGCAAAAGACTAACAACCTGAGGGG	421	
DB	380 GTGGCCACAGCGCTCTGTGTGTGTGTGATCGAACCCTGAGCAAAAGACTAACAACCTGAGGGG	321	
QY	422 ATTCATCTTGAAGAGAGAGAGATGCTGTGCGCTTTGAGACTCCAAAGGCTTGCTTT	481	
DB	320 ATTCATCTTGAAGAGAGAGAGATGCTGTGCGCTTTGAGACTCCAAAGGCTTGCTTT	261	
QY	482 ATTATTTGTCTGTTTACTTTTGGAAATTCCTACATTAAGATTAATTTGTAAATATG	541	
DB	260 ATTATTTGTCTGTTTACTTTTGGAAATTCCTACATTAAGATTAATTTGTAAATATG	201	
QY	542 GCCTTTCCTACCTCGTGTGTGTGTGTGATCGAATCATAGAGAGACGAGAACACAG	601	
DB	200 GCCTTTCCTACCTCGTGTGTGTGTGTGATCGAATCATAGAGAGACGAGAACACAG	141	
QY	602 AAATGATCTTTGTTTATCTGATCCACGACT--GGAACATTTGTTACAGAGAAACA	658	
DB	140 AAATGATCTTTGTTTATCTGATCCACGACT--GGAACATTTGTTACAGAGAAACA	81	
QY	659 TTGTTTGTGTTATGCTTGAAGGTTAAAAATGATTAACGAATTTACAGTAACATA	718	

Db	80	TTGTTTGCTTAAGCTTGAAGGCTTAATAAATAGATAAACGAATGTTATAGTAAACAATA	21
Qy	719	AAATGCATGAAAAAGCCGA	737
Db	20	AAATGCATGAAAAAGCCGA	2
RESULT 40			
AM555800/c			
LOCUS			
DEFINITION	10259P11.3 N1A Mouse Newborn Ovary cDNA library Mus musculus cDNA	651 bp	RNA linear EST 31-AUG-2000
ACCESSION	AM555800		
VERSION	10259P11.3		
KEYWORDS	clone L0259P11 3', mRNA sequence.		
SOURCE	AM555800.1	GI:7201223	
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.		
AUTHORS	Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagatani,R., Doi,H., Wood,W.H., Ili, Becker,K.G. and Ko,M.S.H.		
TITLE	Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (16)	9127-9132 (2000)
MEDLINE	20381248		
PUBMED	10922068		
COMMENT	Contact: George J. Kargul Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: chta@igsun.grc.nia.nih.gov Plate: L0259 row: F column: 11 Seq primer: -21M13 Forward High quality sequence stop: 651 POLYA=yes.		
FEATURES			
source	Location/Qualifiers		
	1..651		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6J"		
	/db_xref="taxon:L0259P11-3"		
	/db_xref="taxon:10930"		
	/clone="L0259P11"		
	/sex="female"		
	/dex_stage="Newborn Ovary"		
	/lab_host="DH10B"		
	/clone_id="N1A Mouse Newborn Ovary cDNA library"		
	/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site 1: SalI; Site 2: NotI; Total RNAs were extracted from 7 Newborn Ovary. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor from Gibco/BRL)		
	[5'-pGACTAGTCTAGATCGCAGCGCGCCCTTTT-3'] from 2.56ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Yulan Piao."		
BASE COUNT	222 a	133 c	195 t
ORIGIN			





DEFINITION UT-M-EMO-bv1-p-18-0-UI.r1 NIH\_BMAP\_EMO Mus musculus cDNA clone  
IMAGE:5692673 5', mRNA sequence.  
ACCESSION BM946344  
VERSION BM946344.1 GI:19429929  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
1 (bases 1 to 624)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
1..624  
location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5692673"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 18.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1lb="NIH\_BMAP\_EMO"  
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 150 a 134 c 180 g 158 t 2 others  
ORIGIN

Query Match 51.2%; Score 584; DB 12; Length 624;  
Best Local Similarity 99.5%; Pred. No. 1.3e-114;  
Matches 606; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

27 TGAAGAGCGGCGAGGACCTGCGCTCTTCTTGGCACTC-CGGAGCGCAGGCTCCAG 85  
Db 1 TGAAGAGCGGCGAGGAA-CTGCGCTCTTCTTGGCACTCNCGGAGCGCGCTCCAG 59  
Qy 86 TCGGAGGCGCACAATGTTCTCTCAAGAGTGAACGGGTAGGCATGTGAGCTGG 145  
Db 60 TCGGAGGCGCACAATGTTCTCTCAAGAGTGAACGGGTAGGCATGTGAGCTGG 119  
Qy 146 GACGTTGAGTGATACCTGTCGATCGACGGGTCCAGGTGATGATGATGCTGCTTGA 205  
Db 120 GACGTTGAGTGATACCTGTCGATCGACGGGTCCAGGTGATGATGATGCTGCTTGA 179  
Qy 206 TGTCAAGTGAAGAACAGAGAGAGAGCTGTGTGTGTGTGGGAGAGATGAACATTC 265  
Db 180 TGTCAAGTGAAGAACAGAGAGAGAGCTGTGTGTGTGTGGGAGAGATGAACATTC 239

Qy 266 TTCCACAACTGCTGCATGTCCTGTGGTGAAACGAAACAATCGTGGCCCTGNGCCAG 325  
Db 240 TTCCACAACTGCTGCATGTCCTGTGGTGAAACGAAACAATCGTGGCCCTGNGCCAG 299  
Qy 326 CAGACTGGGTAGTCCAAAGAAATCGCAAAATAGAGAGTGGCCAGCGCTCTGTGTGG 385  
Db 300 CAGACTGGGTAGTCCAAAGAAATCGCAAAATAGAGAGTGGCCAGCGCTCTGTGTGG 359  
Qy 386 TTGCTGACCTGGACAAAGACTAAACACTGACGGGGATTCATCTTTGAGAGAGAGAT 445  
Db 360 TTGCTGACCTGGACAAAGACTAAACACTGACGGGGATTCATCTTTGAGAGAGAGAT 419  
Qy 446 GCTGTGGCTTTGAGACTCACCAGAGGCTGCTTATTAATTTGCTGTAGTTTGG 505  
Db 420 GCTGTGGCTTTGAGACTCACCAGAGGCTGCTTATTAATTTGCTGTAGTTTGG 479  
Qy 506 GAATTCCTACATTAAGTAAATTTGTTAAATAATGGCTTCTTACCTTGTGTGTGT 565  
Db 480 GAATTCCTACATTAAGTAAATTTGTTAAATAATGGCTTCTTACCTTGTGTGTGT 539  
Qy 566 GTGTGATGCAATGATAGAGAGCGGAGAACACCAAAATGATTTGTTATCTGTAC 625  
Db 540 GTGTGATGCAATGATAGAGAGCGGAGAACACCAAAATGATTTGTTATCTGTAC 599  
Qy 626 CCAGCACTG 634  
Db 600 CCAGCACTG 608

RESULT 44  
CB318426 757 bp mRNA linear EST 04-MAR-2003  
LOCUS CB318426  
DEFINITION AGNCOURT 12396411 NIH MGC 165 Mus musculus cDNA clone  
IMAGE:30283955 5', mRNA sequence.  
CB318426  
VERSION CB318426.1 GI:28842661  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
1 (bases 1 to 757)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. Leslie L. Heckert  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: NDCM47 row: d column: 12  
High quality sequence stop: 500.  
location/Qualifiers  
1..757  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30283955"  
/tissue\_type="Primary cultures of Sertoli cells"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_1lb="NIH MGC 165"  
/note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI (ggccatcatggcc); Site 2: SfiI (ggcgccctcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTGAGAGCGCGGCGGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.6-3.5 kb). 15/15 colonies contained inserts by

FEATURES  
source



QY		598 CCGAGAAATGATCTTTGTATTATCTGACCCAC	629
Dd		604 CCGAGAAATGATCTTTGTATTATCTGACCCAC	635
RESULT	46		
BP023423		575 bp	mRNA
LOCUS		U08A05.Y1 Soares thymus 2NBMT Mus musculus cDNA clone	
DEFINITION		IMAGE:3470864.5 similar to TR:09WTZ1 Q9WTZ1 ZINC RING FINGER	
ACCESSION		BF023423	
VERSION		BF023423.1	GI:10754756
KEYWORDS		EST.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus,	
TITLE		1. (bases 1 to 575) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL		Unpublished	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nhl.gov This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:1390224 Seq primer: -40RP from Gibco High quality sequence stop: 480. Location/Qualifiers	
FEATURES	source	1..575	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:3470864"	
		/sex="male"	
		/tissue_type="Thymus"	
		/dev_stage="4 weeks"	
		/lab_host="DH10B"	
		/clone_1lb="Soares thymus 2NBMT"	
		/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTTCAACTCATGTAAGTCGGAGCGCGGTTTTTTTTTTTTTTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bernaldo."	
BASE COUNT	.139 a	127 c	170 g 139 t
ORIGIN			
Query Match	50.4%; Score 575; DB 10; Length 575;		
Best Local Similarity	100.0%; Pred. No. 1.1e-112;		
Matches	575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Dd		28 GGAGAGCGCGGAGGAACCTGCGTCTTTCTTGCGACTCCGGAGGCGAGGCCTCCAAGTC	87
OY		1 GGAGAGCGCGGAGGAACCTGCGTCTTTCTTGCGACTCCGGAGGCGAGGCCTCCAAGTC	60
Dd		88 GGGAGGCGACAAGAATGTTCTCTCTCAAGAAGTGAAGCGCGTAGCCATGTGAGAGCTGGGA	147
OY		61 GGGAGGCGACAAGAATGTTCTCTCTCAAGAABTGAAGCGCGTAGCCATGTGAGAGCTGGGA	120
Dd		148 CGTTGAGTGCATACCTGTGCCATCTGCAAGGCTCAGAGTCAAGTGAATGATGCCCTGCGATG	207
OY		121 CGTTGAGTGCATACCTGTGCCATCTGCAAGGCTCAGAGTGAATGATGCCCTGCGATG	180
Dd		208 TCACAGTGAACAACAAGCAAGAGGAGCTGTGTGTGGTCTGGGAGAGTGAACCATTCCTT	267

Db	181	TCGAGCTGAAACCAAGCAAGAGACTGTGTGTGTGTGTGGGAGGTGAACCATTCCTT	240
QY	268	CCACAACTGCTGCATGTCCCTGTGGGTGAACAGAACAATGCTGCGCTCTGTGCCAGA	327
Db	241	CCACAACTGCTGCATGTCTCTGTGGGTGAACAGAACATGCTGCGCTCTGTGCCAGA	300
QY	328	GGACTGGGTAGTCCAAAGATCGGCAAATGAGAGTGTCGCCAGCGCCTCTGGTGTGTT	387
Db	301	GGACTGGGTAGTCCAAAGATCGGCAAATGAGAGTGTCGCCAGCGCCTCTGGTGTGTT	360
QY	388	GCTGACCCCTGAGCAAGACCTTAACACTGCAGGGGATTCATCCTTGAGAGAGAGATGC	447
Db	361	GCTGACCCCTGAGCAAGACCTTAACACTGCAGGGGATTCATCCTTGAGAGAGAGATGC	420
QY	448	TGTGGCGCTTTGAGACCTCACCAAGGCGTTCCTTAATTAATTTGTTAGTTTTGGGA	507
Db	421	TGTGGCGCTTTGAGACCTCACCAAGGCGTTCCTTAATTAATTTGTTAGTTTTGGGA	480
QY	508	AATTCTCTACAAATGAATATTGTTAAAAATGGCCTTTCCTAACCCTGTGGTGTGTGT	567
Db	481	AATTCTCTACAAATGAATATTGTTAAAAATGGCCTTTCCTAACCCTGTGGTGTGTGT	540
QY	568	GTGATACGAATGCATAGAGAGCGAGAACACCCAGA	602
Db	541	GTGATACGAATGCATAGAGAGCGAGAACACCCAGA	575

FEATURES	776 bp	mRNA	linear	EST	03-APR-2003
RESULT 47					
CB600846					
LOCUS					
DEFINITION	Agencourt 13037376 NIH MGC 176 Mus musculus cDNA clone				
ACCESSION	CB600846				
VERSION	GI:29518702				
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 776)				
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished				
	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgrabbs@mail.nih.gov">cgrabbs@mail.nih.gov</a> Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Michael Brownstein Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> plate: NDCM99 row: a column: 10 High quality sequence stop: 462.				
	Location/Qualifiers				

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30313065"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_176"
/name="Organ: Kidney; Vector: pDMR-LIB; Site 1: SfiI
(ggcgcctcggcgc) / Site 2: SfiI (ggcgcctcggcgc) / cDNA made
by oligo-dt priming and directionally cloned, 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGTGTCATCAACGACGATGTCATTACGGCCGG-3' and
5'-ATTTCAGAGCGCGAGCGCGCCGACATC-dT(30)NM-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction library created in the laboratory of M.
Brownstein (NIH, NIH). Note: this is a NIH_MGC Library.

```

BASE COUNT 190 a 169 c 213 g 204 t  
ORIGIN

Query Match 50.2%; Score 572.8; DB 14; Length 776;  
Best Local Similarity 98.5%; Pred. No. 3.2e-112;  
Matches 599; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

14 GCCATGCGCCGACCTGAGAGAGCGGAGGAAACCTGCGCTCTTCTTCCGACTCCGGAGC 73  
4 GCCATGCGCCGACCTGAGAGAGCGGAGGAAACCTGCGCTCTTCTTCCGACTCCGGAGC 63  
74 GGAAGCTCCAAAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGAACCGCGTAGCC 133  
64 GCAGGCTCCAAAGTGGGAGGCGCAAGATGTTCTCTCAAGAGTGAACCGCGTAGCC 123  
134 ATGTGAGCTGGAGACCTTGAAGTCCATCTGTCGCACTGCGAGGTCAGGTGATGAT 193  
124 ATGTGAGCTGGAGACCTTGAAGTCCATCTGTCGCACTGCGAGGTCAGGTGATGAT 183  
194 GCCTGCTTCGATGTCAAGTGAAGAAACAAAGCAAGAGACTGTGTGCTGGGAGAG 253  
184 GCCTGCTTCGATGTCAAGTGAAGAAACAAAGCAAGAGACTGTGTGCTGGGAGAG 243  
254 TGTAAACATTCCTTCCAAACCTGCTCATGTCCTGCTGGGTGAACAGAACATCGCTC 313  
244 TGTAAACATTCCTTCCAAACCTGCTCATGTCCTGCTGGGTGAACAGAACATCGCTC 303  
314 CCTCTGTCCAGAGACCTGCTGATGTCACAAAGATGCGCAATGAGAGTGGCCCGAGCG 373  
304 CCTCTGTCCAGAGACCTGCTGATGTCACAAAGATGCGCAATGAGAGTGGCCCGAGCG 363  
374 CTCCTGTGTGTGTGCTGACCCCTGACAAAGACTAAACACTGAGGGGATTCATCTTGA 433  
364 CTCCTGTGTGTGTGCTGACCCCTGACAAAGACTAAACACTGAGGGGATTCATCTTGA 423  
434 GAGAGAGAGATGCTGTGCGCCCTTTGAGACTCAACAAAGCTTGTATTAATTTGCT 493  
424 GAGAGAGAGATGCTGTGCGCCCTTTGAGACTCAACAAAGCTTGTATTAATTTGCT 483  
494 GTTTAGTTTGGGAATTTCTTCAATTAAGATTAATTTG- TTTAAATGCGCTTCTTAC 552  
484 GTTTAGTTTGGGAATTTCTTCAATTAAGATTAATTTGTTTAAATGCGCTTCTTAC 543  
553 CTCCTGTGTGTGTGCTGATGACAAATGACATA-GAAGAGAGAGACCCGAAATATGATC 611  
544 CTCCTGTGTGTGTGCTGATGACAAATGACATA-GAAGAGAGAGACCCGAAATATGATC 603  
612 TTTGTTAT 619  
604 TTTGTTT 611

RESULT 48  
BI153818 735 bp mRNA linear EST 05-JUL-2001  
LOCUS 602870803F1 NCI\_GCAP\_Mam2 Mus musculus cDNA clone IMAGE:5002729 5',  
DEFINITION  
BI153818 mRNA sequence.  
VERSION  
BI153818.1 GI:14613819  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
NIH-MGC http://mgc.nci.nih.gov/  
1 (bases 1 to 735)  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1038 row: e column: 02  
High quality sequence stop: 622.  
Location/Qualifiers

FEATURES  
source

1..735  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5002729"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_1ib="NCI\_GCAP\_Mam2"  
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 173 a 185 c 201 g 176 t  
ORIGIN

Query Match 49.9%; Score 568.8; DB 12; Length 735;  
Best Local Similarity 96.9%; Pred. No. 2.3e-111;  
Matches 624; Conservative 0; Mismatches 12; Indels 8; Gaps 4;

20 GCCGACGTGAGAGACCGGAGGAAACCTGCTCTTCTTCCGACTCCGGAGCGGAGC 79  
1 GCCGACGTGAGAGACCGGAGGAAACCTGCTCTTCTTCCGACTCCGGAGCGGAGC 60  
80 TCCAA-GTGGGAGGCGCAACATGTTCTCTCAAGAGTGAACCGCGTAGCATGTC 138  
61 TCCAAATGTCGAGGCGCAACATGTTCTCTCAAGAGTGAACCGCGTAGCATGTC 120  
139 GAGCTGGGACGTTGAGTGGATACCTGTCATCTGCAAGGTCGAGGTGATGATGCTG 198  
121 GAGCTGGGACGTTGAGTGGATACCTGTCATCTGCAAGGTCGAGGTGATGATGCTG 180  
199 CCTTGATGTCAGCTGTAACAAACAAAGAGAGACTGTGTTGTGCTGGGAGAGGTGA 258  
181 CCTTGATGTCAGCTGTAACAAACAAAGAGAGACTGTGTTGTGCTGGGAGAGGTGA 240  
258 CCATTCCTTCCCAACTGCTGATGTCCTGCTGGGTGAACCAACATGCTGCTCT 318  
241 CCATTCCTTCCCAACTGCTGATGTCCTGCTGGGTGAACCAACATGCTGCTCT 300  
319 GTGCCAGAGACGTGGGTAGTCCAAAGATGGGCAATGAGAGGTGGCCAGCGCTCT 378  
301 GTGCCAGAGACGTGGGTAGTCCAAAGATGGGCAATGAGAGGTGGCCAGCGCTCT 360  
379 GGTGTGTTGCTGACCTGAGCAACAAAGACTTAACATGCGAGGGATTCATCTTGA--GAG 436  
361 GGTGTGTTGCTGACCTGAGCAACAAAGACTTAACATGCGAGGGATTCATCTTGA 420  
437 AGAGAGATGCTGTGCGCTTTGAGACTCACCAAGGCTTCTTTATTAATTTGCTGT 496  
421 GCAGAGATGCTGTGCGCTTTGAGACTCACCAAGGCTTCTTTATTAATTTGCTGT 480  
497 TAGTTTGGGAATTTCTTCAATTAAGATA-ATTGTTAAATGAGGCTTCTTCAACCTC 555  
481 TAGTTTGGGAATTTCTTCAATTAAGATAATTTGTTAAATGAGGCTTCTTCAACCTC 540  
556 TGTGTGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615  
541 TGTGTGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
616 TTAATCTGTAACCAAGCT---GAAACATGTTCTTCAAGAGA 655  
601 TTAATCTGTAACCAAGCTGTTGAGACATGTTGCTCAAGAGA 644



RESULT 49	BF720947	568 bp	mRNA	linear	EST 03-JAN-2001
LOCUS	BF720947				
DEFINITION	BF720947	568 bp	mRNA	linear	EST 03-JAN-2001
ACCESSION	BF720947				
VERSION	BF720947.1	GI:12021877			
KEYWORDS					
ORGANISM					
SOURCE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
source					
BASE COUNT					
ORIGIN					
Query Match					
Best Local Similarity					
Matches					
44	CCCTCCGCTCTTCTCTGCACTCCGGAGCGCAGGCTCCAAGTCGGGAGCGCAAAAGATG	49.7%	Score 567;	DB 10;	Length 568;
1	CCCTCGCTCTTCTCTGCACTCCGGAGCGCAGGCTCCAAGTCGGGAGCGCAAAAGATG	99.8%	Pred. No. 5.5e-111;		
104	TTCTCTCTCAAGAGTGAACCGCGGTGACCTGTGAGCTGGAGCTGGAGCGTTGAGTGCATAC	0;	Mismatches 1;	Indels 0;	Gaps 0;
61	TTCTCTCTCAAGAGTGAACCGCGGTGACCTGTGAGCTGGAGCGTTGAGTGCATAC				
164	TGTGCACTTGAAGGCTCCAGGTATGATGATCGTCTCGATGTCAAGCTGAACCAAG				
121	TGTGCACTTGAAGGCTCCAGGTATGATGATCGTCTCGATGTCAAGCTGAACCAAG				
224	CAAGAGACTGTGTTGTGTGCTGGGAGAGTGAACCTTCTCTTCCCAACTGCTGCATG				
181	CAAGAGACTGTGTTGTGTGCTGGGAGAGTGAACCTTCTCTTCCCAACTGCTGCATG				
284	TCCCTGTGGTGAACAGAAACAATCGCTCCCTCTGTGTCAGACAGAGACTGGGTATCCAA				
241	TCCCTGTGGTGAACAGAAACAATCGCTCCCTCTGTGTCAGACAGAGACTGGGTATCCAA				

QY		344	AGAAATCGGCAATGAGAGTGGCCACGCGCTCCTGTGTGTTCGTGCAGACCCTTGACGAACA	403
Db		301	AGAATCGGCAATGAGAGTGGCCACGCGCTCCTGTGTGTGTGTGCAGACCCTTGACGAACA	360
QY		404	GACTTAACACTGCGAGGGGATTCACTCCTTGAGAGAGAGAGATGCTGTGCCCTTTGAGAC	463
Db		361	GACTTAACACTGCGAGGGGATTCACTCCTTGAGAGAGAGAGATGCTGTGCCCTTTGAGAC	420
QY		464	TCAACCAAGGCTGTGCTTTAATTTGTCTGTGTAGTTTTGGGAAATTCCTCAATTA	523
Db		421	TCAACCAAGGCTGTGCTTTAATTTGTCTGTGTAGTTTTGGGAAATTCCTCAATTA	480
QY		524	GATAATTTGTAAAAAATGGCCCTTCCTPACCTCTGGTGTGTGTGTGATAGAAATGATA	583
Db		481	GATAATTTGTAAAAAATGGCCCTTCCTPACCTCTGGTGTGTGTGTGATAGAAATGATA	540
QY		584	GAGAGCGAGAACACCCAGAAAATGATCT	611
Db		541	NAAAGCGAGAACACCCAGAAAATGATCT	568
RESULT	50			
LOCUS	AA230335	577 bp	mRNA	linear EST 26-FEB-1997
DEFINITION	mm04c09.1 Soares mouse JNM612.5 Mus musculus cDNA clone			
ACCESSION	U06697.1 similar to WP:ZK287.5 CE0614 ; mRNA sequence.			
VERSION	AA230335			
KEYWORDS	AA230335.1 GI:1852740			
SOURCE	EST.			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 577)			
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubnue,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheinberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
	The WashU-HMNI Mouse EST Project			
TITLE	Unpublished			
JOURNAL	Contract: Marra/Mouse EST Project			
COMMENT	WashU-HMNI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGR Consortium (info@image.jlni.gov) for further information. MG1:409416 Seq primer: -28mJ rev2 ET from Amersham High quality sequence rev2: 498. Location/Qualifiers 1..577 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGR:669712" /sex="unknown" /tissue_type="fetus" /dev_stage="12.5dpc total fetus" /lab_host="DH10B" /clone_1lb="Soares mouse JNM612.5" /note="Organ: whole fetus; Vector: pUT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGCTGGAGCGCGCTATTATTTTATTTTATTTT 3'] , on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into			



the Not I and Eco RI sites of the modified pT7T3 vector.  
Library went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT

142 a 124 c 164 g 147 t

ORIGIN

Query Match

48.8%; Score 556.8; DB 9; Length 577;

Best Local Similarity 99.5%; Pred. No. 8.5e-109;

Matches 569; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
QY 63 ACTCCGGAGGCGCAGGCTCCAGTCCGAGGCGCAAGATGTTCTCTCAGAGAGTGA 122
Db 6 ACTCCGGAGGCGCAGGCTCCAGTCCGAGGCGCAAGATGTTCTCTCAGAGAGTGA 64
QY 123 ACCGGTAGCCATGTGGAGCTGGAGCTGATGCGATCCTGCACTCGAGGGTCC 182
Db 65 ACCGGTAGCCATGTGGAGCTGGAGCTGATGCGATCCTGCACTCGAGGGTCC 124
QY 183 AGGTATGATGCTGCTGCTGATGTCAGCTGAAAAACAAGAGAGACTGTGTGTGG 242
Db 125 AGGTATGATGCTGCTGCTGATGTCAGCTGAAAAACAAGAGAGACTGTGTGTGG 184
QY 243 TCTGGGAGAGTAACTTCTTCCAACTGCTGATGCTGCTGCTGCTGCTGCTGCT 302
Db 185 TCTGGGAGAGTAACTTCTTCCAACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 244
QY 303 ACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
Db 245 ACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 304
QY 363 TGGCCAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
Db 305 TGGCCAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 364
QY 423 TTCACTTGAAGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482
Db 365 TTCACTTGAAGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 424
QY 483 TTAATTGCTGTTAGTTTGGGAAATCTCAATTAAGTAATTTGTTAAAAATGG 542
Db 425 TTAATTGCTGTTAGTTTGGGAAATCTCAATTAAGTAATTTGTTAAAAATGG 484
QY 543 CCTTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 485 CCTTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544
QY 603 AAATGATCTTTGTTATCTGTACCAAGACTG 634
Db 545 AAATGATCTTTGTTATCTGTACCAAGACTG 576
```

Search completed: November 7, 2003, 15:15:38  
Job time : 2889.08 secs

**THIS PAGE BLANK (USPTO)**



```
RESULT 2
US-10-108-767-7
; Sequence 7, Application US/10108767
; Publication No. US20030104474A1
; GENERAL INFORMATION:
; APPLICANT: Issakani, Sarkiz D.
; APPLICANT: Huang, Jiarng
; APPLICANT: Sheung, Julie
; APPLICANT: Pray, Todd R.
; TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENT
; FILE REFERENCE: A-68613-5/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/108,767
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 09/542,497
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 09/826,312
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 10/091,139
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-767-7

Query Match      100.0%; Score 124; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 6,9e-60;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTCTTATGATGCAAGCTGA 60
        |||||||
DB      141 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTCTTATGATGCAAGCTGA 200
QY      61 AAACAACAAGAGAGCTGTGTGTGCTGGGAGAAATGATATCTTCTTCCAACTG 120
        |||||||
DB      201 AAACAACAAGAGAGCTGTGTGTGCTGGGAGAAATGATATCTTCTTCCAACTG 260
QY      121 CTGC 124
        ||||
DB      261 CTGC 264

RESULT 3
US-10-152-156-7
; Sequence 7, Application US/10152156
; Publication No. US20030108947A1
; GENERAL INFORMATION:
; APPLICANT: Issakani, Sarkiz D.
; APPLICANT: Huang, Jiarng
; APPLICANT: Sheung, Julie
; APPLICANT: Pray, Todd R.
; TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENT
; FILE REFERENCE: A-68613-6/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/152,156
; PRIOR FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US 09/542,497
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 09/826,312
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 10/091,174
; PRIOR FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US 10/091,139
; PRIOR FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US 10/109,460
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 10/108,767
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/291,836
```

```
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-152-156-7

Query Match      100.0%; Score 124; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 6,9e-60;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTCTTATGATGCAAGCTGA 60
        |||||||
DB      141 CGATACGTCGCCCATCTGCAGGGTCCAGGTGATGATGCTCTTATGATGCAAGCTGA 200
QY      61 AAACAACAAGAGAGCTGTGTGTGCTGGGAGAAATGATATCTTCTTCCAACTG 120
        |||||||
DB      201 AAACAACAAGAGAGCTGTGTGTGCTGGGAGAAATGATATCTTCTTCCAACTG 260
QY      121 CTGC 124
        ||||
DB      261 CTGC 264

RESULT 4
US-09-796-692-8547
; Sequence 8547, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077_001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8547
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (485)
; OTHER INFORMATION: n=A,T,C or G
```

Query Match 100.0%; Score 124; DB 14; Length 612;

OTHER INFORMATION: n equals a,t,g, or c

100.0%; Score 124; DB 14; Length 612;



```

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7306
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 36, 150, 168, 187, 224, 266, 305, 375, 378, 405, 429, 438,
; LOCATION: 439, 454, 458, 464, 472, 485, 491, 495, 516, 536, 537, 538,
; LOCATION: 544, 545, 548, 550, 558, 559, 565, 567, 570, 578, 581,
; LOCATION: 585, 617, 633, 635, 636, 637, 641, 644, 647, 651, 657
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 672, 674, 680, 690, 692, 694, 695, 699, 704, 705, 707, 709,
; LOCATION: 725, 733, 740, 743, 744, 747, 748, 754, 756, 757, 762, 765,
; LOCATION: 769, 771, 774, 775, 777, 783, 784, 786, 788, 792, 795, 797,
; LOCATION: 798, 801, 802, 805, 809, 812, 813, 814, 815, 817, 818
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 819, 824, 827, 836, 838, 846, 851, 852, 855, 860, 865, 871,
; LOCATION: 876, 882, 893, 901, 903, 904, 905, 908, 909, 910, 913, 915,
; LOCATION: 919, 925, 927, 930, 931
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-7306

Query Match          19.4%; Score 24; DB 14; Length 933;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      101 AATCATTCCTTCCCACTGCTGC 124
          |||||
DB      90 AATCATTCCTTCCCACTGCTGC 113

RESULT 12
US-09-974-026-61/c
; Sequence 61, Application US/09974026
; Publication No. US20030194398A1
; GENERAL INFORMATION:
; APPLICANT: Tamburini, Paul P
; APPLICANT: Davis, Gary
; APPLICANT: Delaria, Katherine A
; APPLICANT: Christopher, Marlor W
; APPLICANT: Daniel, Muller K
; TITLE OF INVENTION: Human Bikunin
; FILE REFERENCE: 96-223-22
; CURRENT APPLICATION NUMBER: US/09/974,026
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/144,428
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: PCT/US97/03894
; PRIOR FILING DATE: 1997-03-10
; PRIOR APPLICATION NUMBER: US 08/725,251
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: US 60/019,793
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: US 60/013,106
; PRIOR FILING DATE: 1996-03-11
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 45
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide used in in vitro mutagenesis in Example 9.
; US-09-974-026-61

Query Match          15.3%; Score 19; DB 12; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 ACGTGGCCATCTGCAGG 23
          |||||
DB      38 ACGTGGCCATCTGCAGG 20

RESULT 13
US-10-027-632-113318/c
; Sequence 113318, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1998-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113318
; LENGTH: 3372
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-113318

Query Match          15.3%; Score 19; DB 12; Length 3372;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      70 AGAGACTGTGTTGGTGC 88
          |||||
DB      1339 AGAGACTGTGTTGGTGC 1321

RESULT 14
US-10-027-632-113318/c
; Sequence 113318, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
```

;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 113318  
;; LENGTH: 3372  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-113318

Query Match 15.3%; Score 19; DB 13; Length 3372;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 AGAGGACTGTGTGTGTC 88  
DB 1339 AGAGGACTGTGTGTGTC 1321

RESULT 15  
US-09-918-995-19059/C  
;; Sequence 19059, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: HySeq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; PRIOR FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 19059  
;; LENGTH: 336  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc.feature  
;; LOCATION: (1)...(336)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-19059

Query Match 14.5%; Score 18; DB 11; Length 336;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108  
DB 135 GGGAGATGTATCATTC 118

RESULT 16  
US-09-998-598-450/C  
;; Sequence 450, Application US/0998598  
;; Patent No. US20020150922A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Stolk, John A.  
;; APPLICANT: Xu, Jiangchun  
;; APPLICANT: Chenuault, Ruth A.  
;; APPLICANT: Meagher, Madelein Joy  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
;; FILE REFERENCE: 210121.561  
;; CURRENT APPLICATION NUMBER: US/09/998,598  
;; PRIOR FILING DATE: 2001-11-16

;; NUMBER OF SEQ ID NOS: 2606  
;; SOFTWARE: Corixa Invention Disclosure Database  
;; SEQ ID NO 450  
;; LENGTH: 541  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-998-598-450

Query Match 14.5%; Score 18; DB 10; Length 541;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108  
DB 154 GGGAGATGTATCATTC 137

RESULT 17  
US-10-027-632-31061  
;; Sequence 31061, Application US/10027632  
;; Publication No. US20030204075A9  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; PRIOR FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 31061  
;; LENGTH: 930  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-31061

Query Match 14.5%; Score 18; DB 12; Length 930;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ATTCTTCACACTGCT 122  
DB 774 ATTCTTCACACTGCT 791

RESULT 18  
US-10-027-632-31061  
;; Sequence 31061, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; PRIOR FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20



;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 31061  
;; LENGTH: 930  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-31061

Query Match 14.5%; Score 18; DB 13; Length 930;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ATTCTTCCCAACTGCT 122  
|||  
DB 774 ATTCTTCCCAACTGCT 791

## RESULT 19

US-10-106-698-799/c  
;; Sequence 799, Application US/10106698  
;; Publication No. US20030109690A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Ruben et al.  
;; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
;; FILE REFERENCE: PA005PI  
;; CURRENT APPLICATION NUMBER: US/10/106,698  
;; CURRENT FILING DATE: 2002-03-27  
;; PRIOR APPLICATION NUMBER: PCT/US00/26524  
;; PRIOR FILING DATE: 2000-09-28  
;; PRIOR APPLICATION NUMBER: US 60/157,137  
;; PRIOR FILING DATE: 1999-09-29  
;; PRIOR APPLICATION NUMBER: US 60/163,280  
;; PRIOR FILING DATE: 1999-11-03  
;; NUMBER OF SEQ ID NOS: 8564  
;; SOFTWARE: PatentIn Ver. 3.0  
;; SEQ ID NO 799  
;; LENGTH: 2630  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (1676)..(1676)  
;; OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-799

Query Match 14.5%; Score 18; DB 14; Length 2630;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTAACTATTC 108  
|||  
DB 1143 GGGAGATGTAACTATTC 1126

## RESULT 20

US-10-027-632-276727/c  
;; Sequence 276727, Application US/10027632  
;; Publication No. US20030204075A9  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; POLYMORPHISMS IN THE HUMAN GENOME  
;; FILE REFERENCE: 108827.129

;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 276727  
;; LENGTH: 252  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-276727

Query Match 13.7%; Score 17; DB 12; Length 252;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCCAGGTGATG 34  
|||  
DB 51 GCAGGGTCCAGGTGATG 35

## RESULT 21

US-10-027-632-276727/c  
;; Sequence 276727, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; POLYMORPHISMS IN THE HUMAN GENOME  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 276727  
;; LENGTH: 252  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-276727

Query Match 13.7%; Score 17; DB 13; Length 252;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCCAGGTGATG 34  
|||  
DB 51 GCAGGGTCCAGGTGATG 35

```
RESULT 22
US-09-864-761-4974
; Sequence 4974, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4974
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031056.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
US-09-864-761-4974

Query Match 13.7%; Score 17; DB 9; Length 355;
Best Local Similarity 100.0%; Pred. No. 19;
```

```
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 GGTCCAGGTGATGATG 38
|||
Db 281 GGTCCAGGTGATGATG 297

RESULT 23
US-09-864-761-21703
; Sequence 21703, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21703
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031056.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
```

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6  
OTHER INFORMATION: NT HIT: AB019786.1, EVALUE 1.20e-02  
OTHER INFORMATION: EST\_HUMAN HIT: AA495851.1, EVALUE 2.20e-01  
US-09-864-761-21703

Query Match 13.7%; Score 17; DB 9; Length 355;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 GGTCCAGGTGATGATG 38  
DB 281 GGTCCAGGTGATGATG 297

RESULT 24  
US-10-066-543-1656  
Sequence 1656, Application US/10066543  
Publication No. US20030087818A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Yugu  
APPLICANT: Pyle, Ruth A.  
APPLICANT: Xu, Jianshun  
APPLICANT: Indrias, Carol Yoseph  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Carter, Darrick  
APPLICANT: Fanger, Gary R.  
APPLICANT: Smith, Carole L.  
APPLICANT: Durham, Margarita  
APPLICANT: Stolk, John A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.563  
CURRENT APPLICATION NUMBER: US/10/066,543  
CURRENT FILING DATE: 2002-01-31  
NUMBER OF SEQ ID NOS: 3417  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1656  
LENGTH: 464  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 211, 453, 454  
OTHER INFORMATION: n = A,T,C or G  
US-10-066-543-1656

Query Match 13.7%; Score 17; DB 14; Length 464;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGCAGGTCAG 28  
DB 433 CCATCTGCAGGTCAG 449

RESULT 25  
US-09-918-995-31855/C  
Sequence 31855, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: HySeq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31855

LENGTH: 500  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(500)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31855

Query Match 13.7%; Score 17; DB 11; Length 500;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGCAGGTCAG 28  
DB 482 CCATCTGCAGGTCAG 466

RESULT 26  
US-10-027-632-145248/C  
Sequence 145248, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 145248  
LENGTH: 712  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-145248

Query Match 13.7%; Score 17; DB 12; Length 712;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ATCTGCAGGTCAGCT 30  
DB 689 ATCTGCAGGTCAGCT 673

RESULT 27  
US-10-027-632-145248/C  
Sequence 145248, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676

```

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145248
; LENGTH: 712
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-145248
```

```
Query Match      13.7%; Score 17; DB 13; Length 712;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      14 ATCTGAGGTCACAGT 30
      |||||
Db      689 ATCTGAGGTCACAGT 673
```

```

RESULT 28
US-10-027-632-162982
; Sequence 162982, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162982
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-162982
```

```
Query Match      13.7%; Score 17; DB 12; Length 833;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      50 TGTCAAGCTGAACAA 66
      |||||
Db      194 TGTCAAGCTGAACAA 210
```

```

RESULT 29
US-10-027-632-162983
; Sequence 162983, Application US/10027632
```

```

; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162983
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-162983
```

```
Query Match      13.7%; Score 17; DB 12; Length 833;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      50 TGTCAAGCTGAACAA 66
      |||||
Db      194 TGTCAAGCTGAACAA 210
```

```

RESULT 30
US-10-027-632-162982
; Sequence 162982, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162982
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-162982
```

```
Query Match      13.7%; Score 17; DB 13; Length 833;
Best Local Similarity 100.0%; Pred. No. 19;
```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 50 TGTCAAGCTGAACAA 66  
|||||  
Db 194 TGTCAAGCTGAACAA 210

RESULT 31  
US-10-027-632-162983  
; Sequence 162983, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 162983  
; LENGTH: 833  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-162983

Query Match 13.7%; Score 17; DB 13; Length 833;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 50 TGTCAAGCTGAACAA 66  
|||||  
Db 194 TGTCAAGCTGAACAA 210

RESULT 32  
US-10-027-632-255720  
; Sequence 255720, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 255720  
; LENGTH: 1781  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-255720

Query Match 13.7%; Score 17; DB 12; Length 1781;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 TTCCTTCACACTGCT 122  
|||||  
Db 1695 TTCCTTCACACTGCT 1711

RESULT 33  
US-10-027-632-255720  
; Sequence 255720, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 255720  
; LENGTH: 1781  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-255720

Query Match 13.7%; Score 17; DB 13; Length 1781;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 TTCCTTCACACTGCT 122  
|||||  
Db 1695 TTCCTTCACACTGCT 1711

RESULT 34  
US-10-281-904-3/c  
; Sequence 3, Application US/10281904  
; Publication No. US20030119036A1  
; GENERAL INFORMATION:  
; APPLICANT: Chun, Miyoung  
; TITLE OF INVENTION: METHODS OF USING 48149, A HUMAN  
; FILE REFERENCE: MP101-174P1M  
; CURRENT APPLICATION NUMBER: US/10/281,904  
; CURRENT FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: 60/335,084  
; PRIOR FILING DATE: 2001-10-31  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 3
; LENGTH: 2904
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-281-904-3

Query Match
Best Local Similarity 13.7%; Score 17; DB 14; Length 2904;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CCATCTGCAGGGTCCAG 28
Db 1642 CCATCTGCAGGGTCCAG 1626

RESULT 35
US-09-880-107-2295/c
; Sequence 2295, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darcí T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2295
; LENGTH: 3494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M22324
US-09-880-107-2295

Query Match
Best Local Similarity 13.7%; Score 17; DB 10; Length 3494;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CCATCTGCAGGGTCCAG 28
Db 1762 CCATCTGCAGGGTCCAG 1746

RESULT 36
US-10-205-823-31/c
; Sequence 31, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatc, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT FILING DATE: 2002-07-25
; CURRENT APPLICATION NUMBER: US/10/205,823
```

```

; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 3494
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-31

Query Match
Best Local Similarity 13.7%; Score 17; DB 14; Length 3494;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CCATCTGCAGGGTCCAG 28
Db 1762 CCATCTGCAGGGTCCAG 1746

RESULT 37
US-10-264-374-200/c
; Sequence 200, Application US/10264374
; Publication No. US20030113320A1
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using
; FILE REFERENCE: P-LJ 3203
; CURRENT APPLICATION NUMBER: US/10/264,374
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: US/09/139,802
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 08/926,914
; PRIOR FILING DATE: 1997-09-10
; PRIOR APPLICATION NUMBER: 08/710,067
; PRIOR FILING DATE: 1996-09-10
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 200
; LENGTH: 3494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(3024)
US-10-264-374-200

Query Match
Best Local Similarity 13.7%; Score 17; DB 14; Length 3494;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CCATCTGCAGGGTCCAG 28
Db 1762 CCATCTGCAGGGTCCAG 1746

RESULT 38
US-10-281-904-1/c
; Sequence 1, Application US/10281904
; Publication No. US20030119036A1
; GENERAL INFORMATION:
; APPLICANT: Chun, Miyoung
; TITLE OF INVENTION: METHODS OF USING 48149, A HUMAN
```

```
/ TITLE OF INVENTION: AMINOPEPTIDASE FAMILY MEMBER
/ FILE REFERENCE: MP101-1741RM
/ CURRENT APPLICATION NUMBER: US/10/281,904
/ CURRENT FILING DATE: 2002-10-28
/ PRIOR APPLICATION NUMBER: 60/335,084
/ PRIOR FILING DATE: 2001-10-31
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 3494
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-281-904-1

Query Match      13.7%; Score 17; DB 14; Length 3494;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CCATCTGCAGGGTCCAG 28
DB      1762 CCATCTGCAGGGTCCAG 1746

RESULT 39
US-10-133-013-221/c
/ Sequence 221, Application US/10133013
/ Publication No. US20030166903A1
/ GENERAL INFORMATION:
/ APPLICANT: Astrimoff, Anna
/ APPLICANT: Bandman, Olga
/ APPLICANT: Cocks, Benjamin G.
/ TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE
/ FILE REFERENCE: PA-0049 US
/ CURRENT APPLICATION NUMBER: US/10/133,013
/ CURRENT FILING DATE: 2002-04-25
/ PRIOR APPLICATION NUMBER: 60/287,067
/ PRIOR FILING DATE: 2001-04-27
/ NUMBER OF SEQ ID NOS: 271
/ SOFTWARE: PERL Program
/ SEQ ID NO 221
/ LENGTH: 3681
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030166903A1 1136923.19
US-10-133-013-221

Query Match      13.7%; Score 17; DB 12; Length 3681;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CCATCTGCAGGGTCCAG 28
DB      1960 CCATCTGCAGGGTCCAG 1944

RESULT 40
US-09-981-353-121/c
/ Sequence 121, Application US/09981353
/ Patent No. US20020160382A1
/ GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy W.
/ APPLICANT: Jones, David A.
/ TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
/ FILE REFERENCE: PA-0038 US
/ CURRENT APPLICATION NUMBER: US/09/981,353
/ CURRENT FILING DATE: 2001-10-11
/ NUMBER OF SEQ ID NOS: 194
/ SOFTWARE: PERL Program
/ SEQ ID NO 121
/ LENGTH: 3686
/ TYPE: DNA
```

```
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20020160382A1 1344279CB1
US-09-981-353-121

Query Match      13.7%; Score 17; DB 10; Length 3686;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CCATCTGCAGGGTCCAG 28
DB      1962 CCATCTGCAGGGTCCAG 1946

RESULT 41
US-10-044-090-218/c
/ Sequence 218, Application US/10044090
/ Publication No. US20020137081A1
/ GENERAL INFORMATION:
/ APPLICANT: Olga Bandman
/ TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
/ FILE REFERENCE: PA-0028 US
/ CURRENT APPLICATION NUMBER: US/10/044,090
/ CURRENT FILING DATE: 2002-01-09
/ NUMBER OF SEQ ID NOS: 850
/ SOFTWARE: PERL Program
/ SEQ ID NO 218
/ LENGTH: 3686
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20020137081A1 1344279CB1
US-10-044-090-218

Query Match      13.7%; Score 17; DB 13; Length 3686;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CCATCTGCAGGGTCCAG 28
DB      1962 CCATCTGCAGGGTCCAG 1946

RESULT 42
US-10-084-817-86/c
/ Sequence 86, Application US/10084817
/ Publication No. US20030119009A1
/ GENERAL INFORMATION:
/ APPLICANT: Susan Stuart
/ APPLICANT: Jed G. Nuchtern
/ APPLICANT: Sharon E. Pion
/ APPLICANT: Jason M. Shohet
/ TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
/ FILE REFERENCE: PA-0046 US
/ CURRENT APPLICATION NUMBER: US/10/084,817
/ CURRENT FILING DATE: 2002-02-25
/ PRIOR APPLICATION NUMBER: 60/270,784
/ PRIOR FILING DATE: 2001-02-23
/ NUMBER OF SEQ ID NOS: 365
/ SOFTWARE: PERL Program
/ SEQ ID NO 86
/ LENGTH: 3686
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030119009A1 1344279CB1
US-10-084-817-86

Query Match      13.7%; Score 17; DB 14; Length 3686;
Best Local Similarity 100.0%; Pred. No. 19;
```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGACGGGTCCAG 28  
Db 1962 CCATCTGACGGGTCCAG 1946

## RESULT 43

US-09-925-297-338/c  
; Sequence 338, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 338  
; LENGTH: 3728  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3707)  
; OTHER INFORMATION: n equals a,t,g, or c  
; LOCATION: (3713)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-338

Query Match 13.7%; Score 17; DB 9; Length 3728;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGACGGGTCCAG 28  
Db 1972 CCATCTGACGGGTCCAG 1956

## RESULT 44

US-10-133-013-222/c  
; Sequence 222, Application US/10133013  
; Publication No. US20030166903A1  
; GENERAL INFORMATION:  
; APPLICANT: Astromoff, Anna  
; APPLICANT: Bandman, Olga  
; APPLICANT: Cocks, Benjamin G.  
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE  
; FILE REFERENCE: PA-0049 US  
; CURRENT APPLICATION NUMBER: US/10/133,013  
; CURRENT FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 60/287,067  
; PRIOR FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 271  
; SOFTWARE: PRL Program  
; SEQ ID NO 222  
; LENGTH: 3815  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030166903A1 1136923.18  
US-10-133-013-222

Query Match 13.7%; Score 17; DB 12; Length 3815;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGACGGGTCCAG 28  
Db 1961 CCATCTGACGGGTCCAG 1945

## RESULT 45

US-09-764-891-6245/c  
; Sequence 6245, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 6245  
; LENGTH: 11150  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6245

Query Match 13.7%; Score 17; DB 11; Length 11150;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGACGGGTCCAG 28  
Db 76 CCATCTGACGGGTCCAG 60

## RESULT 46

US-10-205-428-561/c  
; Sequence 561, Application US/10205428  
; Publication No. US20030108907A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA117C1  
; CURRENT APPLICATION NUMBER: US/10/205,428  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1019  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 561  
; LENGTH: 11150  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-205-428-561



Query Match 13.7% Score 17; DB 14; Length 11150;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCTGCAGGCTCCAG 28  
DB 76 CCATCTGCAGGCTCCAG 60

## RESULT 47

US-09-878-574-8350  
; Sequence 8350, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 8350  
; LENGTH: 253  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701101059H1  
US-09-878-574-8350

Query Match 12.9% Score 16; DB 10; Length 253;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 TGAACAACAACAAGAG 73  
DB 115 TGAACAACAACAAGAG 130

## RESULT 48

US-09-682-706-1/C  
; Sequence 1, Application US/09682706  
; Patent No. US20020082409A1  
; GENERAL INFORMATION:  
; APPLICANT: Hsu, Sheau-Yu  
; APPLICANT: Hsueh, Aaron  
; TITLE OF INVENTION: Stresscopins and their ses  
; FILE REFERENCE: STAN210  
; CURRENT APPLICATION NUMBER: US/09/682,706  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: 60/276,615  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/244,128  
; PRIOR FILING DATE: 2000-10-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 339  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-682-706-1

Query Match 12.9% Score 16; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GTGCTGTGGGAGAAAT 98  
DB 116 GTGCTGTGGGAGAAAT 101

## RESULT 49

US-09-983-965-5910/C  
; Sequence 5910, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengping  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 5910  
; LENGTH: 397  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 64-LIB34-071-Q1-E1-H8  
US-09-983-965-5910

Query Match 12.9% Score 16; DB 10; Length 397;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGGCTCCAGGTGATGG 35  
DB 88 AGGCTCCAGGTGATGG 73

## RESULT 50

US-09-919-473-1/C  
; Sequence 1, Application US/09919473  
; Patent No. US20020127221A1  
; GENERAL INFORMATION:  
; APPLICANT: Vale, Wylie Walker Jr.  
; APPLICANT: Lewis, Kathy Ann  
; APPLICANT: Reyes, Teresa Marie  
; APPLICANT: Hogenesch, John Beien  
; APPLICANT: Sawchenko, Paul Emil  
; APPLICANT: Vaughan, Joan Maureen  
; APPLICANT: Rivier, Jean Edouard Frederic  
; APPLICANT: Perrin, Marilyn Heller  
; TITLE OF INVENTION: Urocortin Proteins and Uses Thereof  
; FILE REFERENCE: D6334  
; CURRENT APPLICATION NUMBER: US/09/919,473  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/273,969  
; PRIOR FILING DATE: 2001-03-07  
; NUMBER OF SEQ ID NOS: 13  
; SEQ ID NO 1  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 7..345  
; OTHER INFORMATION: DNA Sequence encoding human  
; OTHER INFORMATION: Urocortin-related peptide (hURP)  
US-09-919-473-1

Query Match 12.9% Score 16; DB 10; Length 399;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GTGCTGTGGGAGAAAT 98  
DB 116 GTGCTGTGGGAGAAAT 101

Fri Nov 7 15:17:33 2003

us-09-509-779-3\_copy\_141\_264.rnpb

Page 16

Db 122 GTGCTCTGGGAGAAAT 107

Search completed: November 7, 2003, 13:55:40  
Job time : 479.215 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 02:39:56 / Search time 126.47 Seconds  
(without alignments)  
2646.715 Million cell updates/sec

Title: US-09-509-779-3\_COPY\_141\_264

Perfect score: 124

Sequence: 1 CGATACGTCGCGCATCTGCA.....ATTCTTCACAACTGCTGC 124

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_19jun03.\*

```
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	124	100.0	342	21	AAA96883	Nucleotide sequenc
2	124	100.0	342	22	AAH43570	ROC2 coding sequen
3	124	100.0	342	24	AAH39674	Human RING finger
4	124	100.0	754	20	AAH87325	Human sensitive to
5	124	100.0	754	20	AAH87330	Human sensitive to
6	124	100.0	754	20	AAH87331	Human sensitive to
7	124	100.0	754	20	AAH87314	Human sensitive to
8	124	100.0	836	22	AAH25860	Human CDNA encodin

9	124	100.0	836	22	AAH26319	Human CDNA encodin
10	124	100.0	836	25	ABX73201	Human novel polynu
11	124	100.0	836	25	ABX73660	Human novel polynu
12	124	100.0	962	21	AAH77504	Human OREF ORF3059
13	121	97.6	754	20	AAH87324	Human sensitive to
14	116	93.5	754	20	AAH87317	Human sensitive to
15	112	90.3	754	20	AAH87323	Human sensitive to
16	107	86.3	754	20	AAH87318	Human sensitive to
17	107	86.3	754	20	AAH87319	Human sensitive to
18	103	83.1	754	20	AAH87322	Human sensitive to
19	103	83.1	754	20	AAH87326	Human sensitive to
20	97	78.2	754	20	AAH87321	Human sensitive to
21	90	72.6	747	20	AAH87315	Human sensitive to
22	89	71.8	1152	22	AAH25847	Human apoptosias as
23	83	66.9	754	20	AAH87320	Human sensitive to
24	76	61.3	754	20	AAH87328	Human sensitive to
25	74	59.7	754	20	AAH87327	Human sensitive to
26	74	59.7	754	20	AAH87329	Human sensitive to
27	43	34.7	224	20	AAH41085	Human secreted pro
28	42	33.9	706	20	AAH87316	Human secreted pro
29	41	33.1	441	21	AAH77493	Human OREF ORF3048
30	35	28.2	60	24	ABH40537	Human spliced tran
31	35	28.2	439	24	ABH21414	Human polynucleoti
32	32	25.8	1140	20	AAH87313	Mouse sensitive to
33	32	19.4	264	22	AAH16209	Human sensitive to
34	24	19.4	596	22	AAH25052	Human breast cance
35	24	19.4	630	22	AAH22239	Human breast cance
36	19	15.3	45	21	AAH70390	Site directed muta
37	18	14.5	18	20	AAH87337	Human sensitive to
38	18	14.5	349	21	AAH29980	Human secreted pro
39	18	14.5	541	24	ABH87139	Human colon cancer
40	18	14.5	2405	15	AAH070827	Protein kinase (CK
41	18	14.5	2405	16	AAH092962	Human HR235-like c
42	18	14.5	2630	22	AAH33733	Human colon cancer
43	17	13.7	239	23	AAH84630	DNA encoding novel
44	17	13.7	355	22	ABA46329	Human breast cell
45	17	13.7	355	22	ABA51431	Human breast cell

## ALIGNMENTS

XX	RESULT 1
XX	ID AAA96883
XX	AAA96883 standard; DNA; 342 BP.
XX	AC AAA96883;
XX	DT 19-FEB-2001 (first entry)
XX	DE Nucleotide sequence of human ring finger protein ROC2.
XX	KW ROC1; ROC2; cullin; ring finger protein; APC1; APC complex; SCF pathway;
XX	KW cullin dependent ubiquitin ligase; CDK inhibitor Sic1 degradation;
XX	KW tumour; ss.
XX	OS Homo sapiens.
XX	Key Location/Qualifiers
XX	FT 1..342
XX	FT CDS
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX	FT
XX</	

31-MAR-2000; 2000OWO-US08592.  
31-MAR-1999; 99US-0127261.  
PR 22-NOV-1999; 99US-0166927.  
XX  
XX (UYNC-) UNIV NORTH CAROLINA.  
XX  
XX  
XX  
PI Xiong Y, Ohta T;  
DR WPI: 2000-647235/62.  
DR P-PSDB; AAB19161.  
XX  
XX  
PT Novel nucleic acid encoding cullin regulating ring finger proteins,  
PT termed as ROC proteins similar to anaphase-promoting complex 11, for  
PI therapeutic and diagnostic use -  
XX  
XX  
PS Claim 18; Fig 2B; 83pp; English.  
XX  
XX  
XX The present sequence encodes a human ROC2 ring finger protein. The  
CC specification also describes human ROC1, ROC1 and ROC2 are similar  
CC to APC11, a subunit of the APC complex. The proteins stimulate cullin  
CC dependent ubiquitin ligase activity. ROC1 functions in vivo as an  
CC essential regulator of CDK inhibitor Sic1 degradation by the SCF  
CC (undefined) pathway. ROC proteins are useful for screening bioactive  
CC agents that interfere with the binding of ROC proteins with cullin  
CC proteins. Pharmaceutical formulations comprising ROC proteins are  
CC useful for diagnostic and therapeutic purposes, preferably for  
XX diagnosing and treating tumours.  
XX  
XX Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other;

XX Issakani SD, Huang J, Sheung J, Pray TR;  
PI WPI: 2001-626445/72.  
DR P-PSDB, AAB47600.  
XX  
XX  
PT Assaying ubiquitin ligase activity for identifying modulators of  
PT ubiquitination, by combining ubiquitin, ubiquitin activating  
PT conjugating enzyme, ubiquitin ligase and measuring amount of ubiquitin  
PT bound to the ligase -  
PS  
PS Example 1; Fig 12A; 98pp; English.  
XX  
XX The sequences given in AAH43568-72 encode proteins which may be used  
CC in the method of the invention for assaying ubiquitin ligase activity.  
CC The method comprises combining under conditions that favour ubiquitin  
CC ligase activity, tag1-ubiquitin, E1 (ubiquitin activating enzyme), E2  
CC (ubiquitin conjugating enzyme) and E3 (ubiquitin ligase) and measuring  
CC the amount of tag1-ubiquitin bound to the E3. The method is useful for  
CC assaying ubiquitin ligase activity and ubiquitination enzyme activity  
CC which is useful for identifying ubiquitination modulator. The method  
CC comprises combining tag1-ubiquitin, the modulator, E1, E2 and tag2-E3  
CC and measuring the amount of tag1-ubiquitin bound to tag2-E3 or combining  
CC tag1-ubiquitin, modulator, E1 and tag3-E2 and measuring the amount of  
CC tag1-ubiquitin bound to tag3-E2. Ubiquitin ligase activity is measured  
CC directly eliminating the need for target proteins and subsequent  
CC analysis such as separating, ligated from unlabeled material in an  
CC SDS-PAGE procedure. This allows multi-well array analysis and high  
CC throughput screening techniques for modulators of ubiquitination  
CC activity. The method also allows the analysis of many different

Query Match	100.0%	Score 124	DB 21	Length 342
Best Local Similarity	100.0%	Pred. No. 1e-55		
Matches 124	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	CGATACGTGCGCCATCTGACGGGTCCAGGTGATGATGCTCTTAAATGTCAGACTGA	60	

Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other; CC measurement of ligated ubiquitin.

QY	1	CGATACCTGGCCATCTGCAGGGSTCCAGGTANAGATGGCTGTCTTAATATCAAGGTGA	60
Db	141	CGATACCTGGCCATCTGCAGGGSTCCAGGTATGATGGCTGTCTTAATATCAAGCTGA	200
QY	61	AAACAACAGAGGACTGTGTGTGGTCTGGGGAGAAATGTATCTATCTCTCCACAATG	120
Db	201	AAACAACAGAGGACTGTGTGTGGTCTGGGGAGAAATGTATCTATCTCTCCACAATG	260
QY	121	CTGC 124	
Db	261	CTGC 264	

QY	DB	Score	Length	Mismatches	Indels	Gaps
1	CGATACGTCGCGCATCTGCAGAGGTC	100.0%	342	0	0	0
141	CGATACGTCGCGCATCTGCAGAGGTC	100.0%	342	0	0	0
61	AAACCAACAGAGGACCTGCTGTTGGTGGGAGAAATATATATCTCTCCCAACTG	100.0%	342	0	0	0

RESULT 2	
AAH43570	
ID	AAH43570 standard; cDNA, 342 bp
XX	
AC	AAH43570;
XX	
DT	07-JUN-2002 (first entry)
XX	
DE	ROC2 coding sequence.
XX	

QY	121	CTGC	124
Db	261	CTGC	264

RESULT 3  
AAD39674  
AAD39674 standard; DNA; 342 BP.

km Assay; ubiquitin ligase; teg1-ubiquitin; E1; E2;  
 kw ubiquitin activating enzyme; ubiquitin conjugating enzyme; E3;  
 xx ubiquitin ligase; ubiquitination modulator; ss.  
 Unidentified.  
 OS

AC	AAD39674;
XX	
DT	22-OCT-2002 (first entry)
XX	
XX	Human RING finger protein, ROC2 DNA.

PN	WO200175145-A2
XX	
PD	11-OCT-2001.
WT	

Ubiquitin ligase; U1; tag1-ubiquitin; E1; ubiquitin-activating enzyme; E2; ubiquitin-conjugating enzyme/ubiquitin carrier protein; E3; human; RING finger protein; gene; ds.

PF 03-APR-2001; 2001WO-US10906.

Homo sapiens.

PR 03-APR-2000; 2000US-0542497.

Key	Location/Qualifiers
-----	---------------------

PA (RIGE-) RIGEL PHARM INC.

1.:342  
/\*tad= a

```
Location/Qualifiers
1..342
/*tag= a
```

FT XX /product= "Human ROC2 protein"  
PN XX US2002042083-A1.  
XX  
PD 11-APR-2002.  
XX  
XX 03-APR-2001; 2001US-0826312.  
PF  
XX 03-APR-2000; 2000US-0542497.  
PR  
XX  
XX (RIGF-) RIGEL PHARM INC.  
PA  
XX Issakani SD, Huang J, Sheung J, Pray TR;  
PI  
XX WPI; 2002-488718/52.  
DR  
XX P-PSDB; AAE24621.  
DR  
PT Assay for ubiquitin ligase activity, useful for identifying modulators,  
PT by measuring binding of labeled ubiquitin to ubiquitin ligase  
XX  
XX  
PS Disclosure; Fig 12A; 56pp; English.  
XX  
XX The invention relates to an assay for ubiquitin ligase (UL) activity  
CC which comprises incubating tag1-ubiquitin, E1 (ubiquitin-activating  
CC enzyme), E2 (ubiquitin-conjugating enzyme/ubiquitin carrier protein)  
CC and E3 (UL) and measuring the amount of tag1-ubiquitin bound to E3. The  
CC method is particularly used to screen for modulators of UL activity. The  
CC present sequence is human RING finger protein, ROC2 DNA.  
XX  
SQ Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other;  
  
Query Match 100.0%; Score 124; DB 24; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1e-55;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCACGCTGA 60  
DB 141 CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCACGCTGA 200  
QY 61 AAACAAACAGAGAGACTGTGTGTGCTGTGGGAGATGATTAATCATTCCTTCCACAACG 120  
DB 201 AAACAAACAGAGAGACTGTGTGTGCTGTGGGAGATGATTAATCATTCCTTCCACAACG 260  
QY 121 CTGC 124  
DB 261 CTGC 264  
  
RESULT 4  
AA87325  
ID AA87325 standard; cDNA; 754 BP.  
XX  
XX AA87325;  
XX  
XX 27-SEP-1999 (first entry)  
XX  
XX Human sensitive to apoptosis (SAG) gene mutant MM9.  
DE  
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnetary; therapy; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH 1..342  
FT CDS /\*tag= a  
FT mutation /\*tag= b  
FT /\*tag= "C99S mutation"  
XX  
XX WO932514-A2.  
PN

XX  
PD 01-JUL-1999.  
XX  
XX 15-DEC-1998; 98WO-US26705.  
PF  
XX 11-SEP-1998; 98US-0099840.  
PR  
XX 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN ) WARNER LAMBERT CO.  
PA  
XX Sun Y;  
PI  
XX WPI; 1999-430152/36.  
DR  
XX P-PSDB; AAY06503.  
DR  
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
XX Claim 15; Page 72-73; 84pp; English.  
XX  
XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM9, which codes for a SAG protein (see AAY06503)  
CC in which the Cys residue at position 99 of the native protein (see  
CC AAY06492) is replaced by a Ser residue owing to a mutation of codon  
CC 99 from TGC to AGC obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is in zinc-ring finger 2 of SAG. Single and  
CC double SAG mutants (see AA87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM9  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.  
CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.  
XX  
SQ Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;  
  
Query Match 100.0%; Score 124; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1e-55;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCACGCTGA 60  
DB 141 CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTGCTTAGATGTCACGCTGA 200  
QY 61 AAACAAACAGAGAGACTGTGTGTGCTGTGGGAGATGATTAATCATTCCTTCCACAACG 120  
DB 201 AAACAAACAGAGAGACTGTGTGTGCTGTGGGAGATGATTAATCATTCCTTCCACAACG 260  
QY 121 CTGC 124  
DB 261 CTGC 264  
  
RESULT 5  
AA87330  
ID AA87330 standard; cDNA; 754 BP.  
XX  
XX AA87330;  
XX  
XX 27-SEP-1999 (first entry)  
XX  
XX Human sensitive to apoptosis (SAG) gene mutant MM14.  
DE  
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnetary; therapy; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX

```

XX Key Location/Qualifiers
FH CDS 1..342
FT mutation /*tag= a
FT mutation replace(295,C)
FT /*tag= b
FT /*note= "C99S mutation"
FT mutation replace(304,C)
FT /*tag= c
FT /*note= "C102S mutation"
XX
XX WO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06508.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 81; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM14, which codes for a SAG protein (see AAY06508)
XX in which the Cys residues at positions 99 and 102 of the native
XX protein (see AAY06492) are replaced by Ser residues owing to
XX mutations of codons 99 and 102 through site-directed mutagenesis of
XX SAG cDNA. These residues are in zinc ring finger 2 of SAG. Single
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. Haem binding was unaffected in MM14, but no
XX oligomerization was observed, indicating that these residues are
XX important for intermolecular disulfide bond formation. SAG is
XX a novel zinc finger protein that promotes cell growth, protects cells
XX from apoptosis, scavenges oxygen radicals and can be used for the
XX reversion of a tumour phenotype. SAG genes, and mutant SAG genes,
XX can be used to protect cells from apoptosis induced by redox
XX reagents. They can also be used for the recombinant production of
XX SAG proteins, which are molecular targets in the development of
XX drugs against neurodegenerative disorders, cancers and muscle
XX dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;
XX
XX Query Match 100.0%; Score 124; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 1e-55; 0; Indels 0; Gaps 0;
XX Matches 124; Conservative 0; Mismatches 0;
XX
XX 1 CGATACGTGGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTATAGATGCAAGCTGA 60
XX 141 CGATACGTGGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTATAGATGCAAGCTGA 200
XX
XX AAAACAACAAGAGAGACTGTGTGTGCTGGGGAGATGTAATCATCTTCCACAACCTG 120
XX
XX 201 AAAACAACAAGAGAGACTGTGTGTGCTGGGGAGATGTAATCATCTTCCACAACCTG 260
XX
XX 121 CTGC 124
XX
XX 261 CTGC 264
XX
XX RESULT 6
XX AAX87331

```

```

ID AAX87331 standard; cDNA; 754 BP.
XX
XX AAX87331;
AC
XX 27-SEP-1999 (first entry)
XX
XX Human sensitive to apoptosis (SAG) gene mutant MM15.
XX
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX vulnereary; therapy; mutant; ds.
XX
XX Homo sapiens.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..342
FT mutation /*tag= a
FT mutation replace(139,C)
FT /*tag= b
FT /*note= "C47S mutation"
XX
XX WO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06509.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 82-83; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM15, which codes for a SAG protein (see AAY06509)
XX in which the Cys residue at position 47 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 47 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is in a GADPH binding site of SAG. Single
XX and double SAG mutants (see AAX87317-31) were made to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. These properties were unaffected by the MM15
XX mutation. SAG is a novel zinc finger protein that promotes
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversion of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
XX Query Match 100.0%; Score 124; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 1e-55; 0; Indels 0; Gaps 0;
XX Matches 124; Conservative 0; Mismatches 0;
XX
XX 1 CGATACGTGGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTATAGATGCAAGCTGA 60
XX
XX 141 CGATACGTGGCCCATCTGCAGGGTCCAGGTGATGATGCTGCTTATAGATGCAAGCTGA 200
XX
XX AAAACAACAAGAGAGACTGTGTGTGCTGGGGAGATGTAATCATCTTCCACAACCTG 120
XX
XX 201 AAAACAACAAGAGAGACTGTGTGTGCTGGGGAGATGTAATCATCTTCCACAACCTG 260
XX
XX 121 CTGC 124
XX
XX 261 CTGC 264
XX
XX RESULT 6
XX AAX87331

```

Db 201 AACCAACAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAACCTG 260  
QY 121 CTGC 124  
Db 261 CTGC 264

RESULT 7  
AA87314  
ID AA87314 standard; cDNA; 754 BP.

XX AC AA87314;

DT 27-SEP-1999 (first entry)

XX DE Human sensitive to apoptosis (SAG) gene.

KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KM neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerable; therapy; ds.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 1..342

FT /\*tag= a

PN WO9932514-A2.

XX 01-JUL-1999.

PF 15-DEC-1998; 98WO-US26705.

PR 11-SEP-1998; 98US-0099840.

PR 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

PI Sun Y;

XX WPI; 1999-430152/36.

DR P-PSDB; AAY06492.

PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis

PS Claim 9; Page 50-51; 84pp; English.

XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) cDNA, which codes for a novel redox-sensitive, hem-binding  
CC protein (see AAY06492) with a zinc RING finger domain that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.  
CC The cDNA was isolated from a HeLa cell cDNA library using mouse  
CC SAG cDNA (see AAX87313) as probe. SAG is highly conserved among  
CC species. Disruption in yeast was shown to be lethal. SAG deletion  
CC mutants (see AAX87315-16) have been identified in human cancer lines,  
CC suggesting a role in carcinogenesis. SAG genes, and mutant SAG  
CC genes, can be used to protect cells from apoptosis induced by redox  
CC reagents. Antisense SAG genes can be used to inhibit the growth of  
CC tumour cells. The SAG genes can also be used for the recombinant  
CC production of the SAG proteins. The SAG proteins can be used to  
CC scavenge oxygen radicals in organisms and to promote wound healing.  
CC Additionally, the SAG genes or their complements can be used to  
CC promote or inhibit the growth of plant cells (all claimed). The SAG  
CC protein is also an ideal molecular target in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy.

XX Sequence 754 BP; 205 A; 155 C; 201 G; 193 T; 0 other;

Query Match 100.0%; Score 124; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1e-55;

Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATAGTGGCGCATCTGCAGAGGTCAGAGTGATGATGCTGCTTATGATGTAAGCTGA 60  
Db 141 CGATAGTGGCGCATCTGCAGAGGTCAGAGTGATGATGATGCTGCTTATGATGTAAGCTGA 200

QY 61 AACCAACAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAACCTG 120  
Db 201 AACCAACAGAGAGACTGTGTGTGTGGGAGATGTAATCATTCCTCCACAACCTG 260

QY 121 CTGC 124  
Db 261 CTGC 264

XX AC AA87314;

DT 27-SEP-1999 (first entry)

XX DE Human sensitive to apoptosis (SAG) gene.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 1..342

FT /\*tag= a

PN WO9932514-A2.

XX 01-JUL-1999.

PF 15-DEC-1998; 98WO-US26705.

PR 11-SEP-1998; 98US-0099840.

PR 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

PI Sun Y;

XX WPI; 1999-430152/36.

DR P-PSDB; AAY06492.

PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis

PS Claim 9; Page 50-51; 84pp; English.

XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) cDNA, which codes for a novel redox-sensitive, hem-binding  
CC protein (see AAY06492) with a zinc RING finger domain that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.  
CC The cDNA was isolated from a HeLa cell cDNA library using mouse  
CC SAG cDNA (see AAX87313) as probe. SAG is highly conserved among  
CC species. Disruption in yeast was shown to be lethal. SAG deletion  
CC mutants (see AAX87315-16) have been identified in human cancer lines,  
CC suggesting a role in carcinogenesis. SAG genes, and mutant SAG  
CC genes, can be used to protect cells from apoptosis induced by redox  
CC reagents. Antisense SAG genes can be used to inhibit the growth of  
CC tumour cells. The SAG genes can also be used for the recombinant  
CC production of the SAG proteins. The SAG proteins can be used to  
CC scavenge oxygen radicals in organisms and to promote wound healing.  
CC Additionally, the SAG genes or their complements can be used to  
CC promote or inhibit the growth of plant cells (all claimed). The SAG  
CC protein is also an ideal molecular target in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy.

XX Sequence 754 BP; 205 A; 155 C; 201 G; 193 T; 0 other;

Query Match 100.0%; Score 124; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1e-55;

```

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-488783/53.
XX P-PSDB; AAU15873.
DR
XX
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX
PS Claim 1; SEQ ID NO 39; 980pp; English.
XX
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

```

Query Match

100.0%; Score 124; DB 22; Length 836;





08-NOV-2000; 2000US-0246611.  
FR 08-NOV-2000; 2000US-0246613.  
FR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
FR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
FR 17-NOV-2000; 2000US-0249218.  
FR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
FR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI, 2001-488783/53.  
XX DR P-PSDB; MAU16332.  
XX  
PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX  
XX  
Claim 1; SEQ ID No 498; 980bp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi. And ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present invention encodes a novel secreted protein of the invention.

	Query Match	100.0%	Score 124;	DB 22;	Length 836;
	Best Local Similarity	100.0%;	Pred. No. 1e-55;		
	Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1 CGATTCGCGCAGCATCTGCAGGGTCCAGGTATGATGATCGCTGTCTTAGATGCAAGCTGA	60			
Dd	209 CGATTCGCGCGCATCTGCAGGGTCCAGGTATGATGATGCCCTGTCTTAATGTCAAGCTGA	268			
Qy	61 AAACAAACAAGAGGACTGTGTTGTGTCTGGGAGATATGTAATCATCTTCACAACTG	120			
Dd	269 AAACAAACAAGAGGACTGTGTTGTGTCTGGGAGAAATGTAATCATCTTCACAACTG	328			
Qy	121 CTGC 124				
Db	329 CTGC 332				
	RESULT 10				
	ABX73201				
ID	ABX73201 standard; DNA; 836 BP.				
XX	ABX73201;				
AC	ABX73201;				
XX					
DT	18-MAR-2003 (first entry)				
XX					
DE	Human novel polynucleotide #29.				
XX					
KW	Human; gene; ds; neural disorder; immune system disorder; renal disorder;				
KM	muscular disorder; respiratory disease; reproductive disorder;				
KW	gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;				
KM	hyperproliferative disorder; inflammatory disease; allergic reaction;				
KW	blood related disorder; cancer; immunosuppressive; antineoplastic;				
KM	cardiovascular; nephrotropic; cyrostatic; antiatherogenic; thrombolytic;				
KW	haemostatic; antiarteriosclerotic.				
XX					
OS	Homo sapiens.				
XX					
PN	US2002132753-A1.				
XX					
PD	19-SEP-2002.				
XX					
PF	17-JAN-2001; 2001US-0764864.				
XX					
PR	31-JAN-2000; 2000US-179065P.				
PR	04-FEB-2000; 2000US-180628P.				
PR	28-JUN-2000; 2000US-214886P.				
PR	07-JUL-2000; 2000US-216647P.				
PR	07-JUL-2000; 2000US-216880P.				
PR	11-JUL-2000; 2000US-217487P.				
PR	11-JUL-2000; 2000US-217496P.				
PR	14-JUL-2000; 2000US-218290P.				
PR	26-JUL-2000; 2000US-220963P.				
PR	26-JUL-2000; 2000US-220964P.				
PR	14-AUG-2000; 2000US-224519P.				
PR	14-AUG-2000; 2000US-224519P.				
PR	14-AUG-2000; 2000US-225267P.				
PR	14-AUG-2000; 2000US-225268P.				
PR	14-AUG-2000; 2000US-225270P.				
PR	14-AUG-2000; 2000US-225447P.				
PR	14-AUG-2000; 2000US-225757P.				
PR	14-AUG-2000; 2000US-225758P.				
PR	22-AUG-2000; 2000US-226868P.				
PR	30-AUG-2000; 2000US-228924P.				
PR	01-SEP-2000; 2000US-229287P.				
PR	01-SEP-2000; 2000US-229343P.				
PR	01-SEP-2000; 2000US-229344P.				
PR	01-SEP-2000; 2000US-229345P.				
PR	05-SEP-2000; 2000US-229509P.				
PR	05-SEP-2000; 2000US-229513P.				
PR	08-SEP-2000; 2000US-231413P.				
PR	21-SEP-2000; 2000US-234223P.				
PR	21-SEP-2000; 2000US-234274P.				
PR	25-SEP-2000; 2000US-234977P.				





```

RESULT 13
ID AAX87324
XX AAX87324 standard; cDNA; 754 BP.
AC AAX87324;
XX
DT 27-SEP-1999 (first entry)
XX
DE Human sensitive to apoptosis (SAG) gene mutant MM8.
XX
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;
KM neurodegenerative disease; muscular dystrophy; wound healing;
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..342
FT /*tag= a
FT replace(262,T)
FT mutation /*tag= b
FT /note= "C88S mutation"
XX
XX MO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98MO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX P-PSDB; AAY06502.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
PT promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 71; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM8, which codes for a SAG protein (see AAY06502)
XX in which the Cys residue at position 88 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 88 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is in zinc-finger 2 of SAG. Single and
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. These properties were unaffected by the MM8
XX mutation. SAG is a novel zinc finger protein that promotes
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversion of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
Query Match 97.6%; Score 121; DB 20; Length 754;
Best Local Similarity 100.0%; Pred. No. 3.9e-54;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGATACGTCGCCATCTGCAGGCTCCAGTGATGATGCCTGCTTAGATGCAAGCTGA 60
|||||

```

```

DB 141 CGATACGTCGCCATCTGCAGGCTCCAGTGATGATGCCTGCTTAGATGCAAGCTGA 200
QY 61 AACAACAAGAGGACTGTTGTGCTGGGGAGATGAATCATCTCCACAACTG 120
|||||
DB 201 AACAACAAGAGGACTGTTGTGCTGGGGAGATGAATCATCTCCACAACTG 260
QY 121 C 121
DB 261 C 261
XX
XX RESULT 14
XX AAX87317
XX ID AAX87317 standard; cDNA; 754 BP.
XX
XX AAX87317;
XX
XX 27-SEP-1999 (first entry)
XX
XX Human sensitive to apoptosis (SAG) gene mutant MM1.
XX
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..342
FT /*tag= a
FT replace(148,T)
FT mutation /*tag= b
FT /note= "C50S mutation"
XX
XX MO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98MO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX P-PSDB; AAY06495.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
PT promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 59; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM1, which codes for a SAG protein (see AAY06495)
XX in which the Cys residue at position 50 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 50 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is at a haem binding site of SAG. Single and
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. MM1 showed reduced haem binding but unaffected
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversion of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX

```

XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;  
SQ Query Match 93.5%; Score 116; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1.7e-51;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 GGGCCATCTGCAGGGTCCAGTGTGATGCTCTTGTAGTCAAGCTGAAACAAAC 68  
DB 149 GGCCATCTGCAGGGTCCAGTGTGATGCTCTTGTAGTCAAGCTGAAACAAAC 208  
QY 69 AAGAGACTGTGTGTGTGTGTGTGGGAGAAATGATCATTCCTCCACAACCTGCTGC 124  
DB 209 AAGAGACTGTGTGTGTGTGTGGGAGAAATGATCATTCCTCCACAACCTGCTGC 264

RESULT 15  
AA87323  
ID AAX87323 standard; cDNA; 754 BP.  
XX  
AC AAX87323;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM7.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnerability; therapy; mutant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..342  
FT FT /\*tag= a  
FT mutation replace(253..255,CAC)  
FT FT /\*tag= b  
FT FT /note= "H85K mutation"  
XX  
FN WO9932514-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 15-DEC-1998; 98WO-US26705.  
XX  
PR 11-SEP-1998; 98US-0099840.  
PR 19-DEC-1997; 97US-0068179.  
XX  
PA (WARN ) WARNER LAMBERT CO.  
XX  
PI Sun Y;  
XX  
DR WPI; 1999-430152/36.  
DR P-PSDB; AAY06501.  
XX  
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
PS Claim 15; Page 69-70; 84pp; English.  
XX  
CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM7, which codes for a SAG protein (see AAY06501)  
CC in which the His residue at position 85 of the native protein (see  
CC AAY06492) is replaced by a Lys residue owing to a mutation of codon  
CC 85 from CAC to AAG obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is in zinc-finger 2 of SAG. Single and  
CC double SAG mutants (see AAX87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM7  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.

CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.  
SQ Sequence 754 BP; 206 A; 153 C; 202 G; 193 T; 0 other;  
QY Query Match 90.3%; Score 112; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 2.2e-49;  
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATCGGCGCCATCTGCAGGGTCCAGTGTGATGCTCTTGTAGTCAAGCTGA 60  
DB 141 CGATCGGCGCCATCTGCAGGGTCCAGTGTGATGCTCTTGTAGTCAAGCTGA 200  
QY 61 AAACAAACAGAGACTGTGTGTGTGTGTGGGAGAAATGATCATTCCTTC 112  
DB 201 AAACAAACAGAGACTGTGTGTGTGTGTGGGAGAAATGATCATTCCTTC 252

RESULT 16  
AAX87318  
ID AAX87318 standard; cDNA; 754 BP.  
XX  
AC AAX87318;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM2.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnerability; therapy; mutant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..342  
FT FT /\*tag= a  
FT mutation replace(157,T)  
FT FT /\*tag= b  
FT FT /note= "C53S mutation"  
XX  
FN WO9932514-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 15-DEC-1998; 98WO-US26705.  
XX  
PR 11-SEP-1998; 98US-0099840.  
PR 19-DEC-1997; 97US-0068179.  
XX  
PA (WARN ) WARNER LAMBERT CO.  
XX  
PI Sun Y;  
XX  
DR WPI; 1999-430152/36.  
DR P-PSDB; AAY06496.  
XX  
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
PS Claim 15; Page 60-61; 84pp; English.  
XX  
CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM2, which codes for a SAG protein (see AAY06496)  
CC in which the Cys residue at position 53 of the native protein (see  
CC AAY06492) is replaced by a Ser residue owing to a mutation of codon  
CC 50 from TGC to AGC obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is at a haem binding site of SAG. Single and  
CC double SAG mutants (see AAX87317-31) were made in order to determine

CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM2  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversal of a tumour phenotype.  
CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.

XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;

Query Match 86.3%; Score 107; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 9.9e-47;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCCAGGTGATGATGCTGCTTATGATGTCAGCTGAAAACAAACAGAGGACT 77  
DB 158 GCAGGGTCCAGGTGATGATGCTGCTTATGATGTCAGCTGAAAACAAACAGAGGACT 217  
QY 78 GTGTGTGCTGGGAGAGATGATTCATTCCTTCCACAACTGCTGC 124  
DB 218 GTGTGTGCTGGGAGAGATGATTCATTCCTTCCACAACTGCTGC 264

## RESULT 17

AAx87319  
ID AAx87319 standard; cDNA; 754 BP.

XX AAX87319;  
XX  
XX 27-SEP-1999 (first entry)

DE Human sensitive to apoptosis (SAG) gene mutant MM3.

XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerary; therapy; mutant; ds.

OS Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers  
XX CDS 1..342  
XX FT /tag= a  
XX FT replace(148,T)  
XX FT /tag= b  
XX FT /note= "C50S mutation"  
XX FT replace(157,T)  
XX FT /tag= c  
XX FT /note= "C53S mutation"

XX WO9932514-A2.

XX 01-JUL-1999.

XX 15-DEC-1998; 98WO-US26705.

XX 11-SEP-1998; 98US-0099840.

XX 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

XX Sun Y;

XX WPI; 1999-430152/36.

XX P-PSDB; AAY06497.

XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
XX promoting cell growth and protecting cells against apoptosis  
XX  
XX Claim 15; Page 62-63; 84pp; English.

XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM3, which codes for a SAG protein (see AAY06497)  
CC in which the Cys residues at positions 50 and 53 of the native  
CC protein (see AAY06492) are replaced by Ser residues owing to  
CC mutations of codons 50 and 53 from TGC to AGC obtained by  
CC site-directed mutagenesis of SAG cDNA. These residues are at a haem  
CC binding site of SAG. Single and double SAG mutants (see AAX87317-31)  
CC were made in order to determine the role of each cysteine residue  
CC of SAG in haem binding and SAG oligomerization. MM3 showed greatly  
CC reduced haem binding but oligomerization was unaffected. SAG is a  
CC novel zinc finger protein that promotes cell growth, protects cells  
CC from apoptosis, scavenges oxygen radicals and can be used for the  
CC reversal of a tumour phenotype. SAG genes, and mutant SAG genes,  
CC can be used to protect cells from apoptosis induced by redox  
CC reagents. They can also be used for the recombinant production of  
CC SAG proteins, which are molecular targets in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
XX dystrophy, and promoting wound healing.

SO Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;

Query Match 86.3%; Score 107; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 9.9e-47;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCCAGGTGATGATGCTGCTTATGATGTCAGCTGAAAACAAACAGAGGACT 77  
DB 158 GCAGGGTCCAGGTGATGATGCTGCTTATGATGTCAGCTGAAAACAAACAGAGGACT 217  
QY 78 GTGTGTGCTGGGAGAGATGATTCATTCCTTCCACAACTGCTGC 124  
DB 218 GTGTGTGCTGGGAGAGATGATTCATTCCTTCCACAACTGCTGC 264

## RESULT 18

AAx87322  
ID AAx87322 standard; cDNA; 754 BP.

XX AAX87322;

XX 27-SEP-1999 (first entry)

DE Human sensitive to apoptosis (SAG) gene mutant MM6.

XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerary; therapy; mutant; ds.

OS Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers  
XX CDS 1..342  
XX FT /tag= a  
XX FT replace(244..246,CAT)  
XX FT /tag= b  
XX FT /note= "H82K mutation"

XX WO9932514-A2.

XX 01-JUL-1999.

XX 15-DEC-1998; 98WO-US26705.

XX 11-SEP-1998; 98US-0099840.

XX 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

XX Sun Y;

XX WPI; 1999-430152/36.





```

XX 01-JUL-1999.
PD 15-DEC-1998; 98MO-US26705.
XX
PF 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
PA (WARN ) WARNER LAMBERT CO.
XX
PI Sun Y;
XX
PI WPI; 1999-430152/36.
DR P-PSDB; AAY06499.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
PT promoting cell growth and protecting cells against apoptosis
XX
PS Claim 15; Page 66; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
CC (SAG) mutant gene MM5, which codes for a SAG protein (see AAY06499)
CC in which the Cys residue at position 80 of the native protein (see
CC AAY06492) is replaced by a Ser residue owing to a mutation of codon
CC 80 from TGC to AGC obtained by site-directed mutagenesis of SAG
CC cDNA. This residue is in zinc-finger finger 1 of SAG. Single and
CC double SAG mutants (see AAX87317-31) were made in order to determine
CC the role of each cysteine residue of SAG in haem binding and SAG
CC oligomerization. These properties were unaffected by the MM5
CC mutation. SAG is a novel zinc finger protein that promotes
CC cell growth, protects cells from apoptosis, scavenges oxygen
CC radicals and can be used for the reversion of a tumour phenotype.
CC SAG genes, and mutant SAG genes, can be used to protect cells from
CC apoptosis induced by redox reagents. They can also be used for the
CC recombinant production of SAG proteins, which are molecular targets
CC in the development of drugs against neurodegenerative disorders,
CC cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
SQ
Query Match 78.2%; Score 97; DB 20; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.9e-41;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGATTGCTGGCCATCTGCAGGGTCCAGGTGATGCTGCTTAGATGTAAGCTGA 60
DB 141 CGATTGCTGGCCATCTGCAGGGTCCAGGTGATGCTGCTTAGATGTAAGCTGA 200
QY 61 AACCAACAAGAGACTGTGTGTGCTGCTGGAGAA 97
DB 201 AACCAACAAGAGACTGTGTGTGCTGCTGGAGAA 237

```

RESULT 21  
AAH87315  
ID AAX87315 standard; cDNA; 747 BP.  
XX  
AC AAX87315;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant 1.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnery; therapy; mutant; colon carcinoma; ds.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT 1..273  
PT CDS  
XX /\*tag= a

```

PN WO9932514-A2.
XX
PD 01-JUL-1999.
XX
PF 15-DEC-1998; 98MO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
PA (WARN ) WARNER LAMBERT CO.
XX
PI Sun Y;
XX
PI WPI; 1999-430152/36.
DR P-PSDB; AAY06493.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
PT promoting cell growth and protecting cells against apoptosis
XX
PS Claim 15; Page 54-55; 84pp; English.
XX
XX This is the nucleotide sequence of deletion mutant 1 of the novel
CC human sensitive to apoptosis gene SAG (see also AAX87314). The cDNA
CC was obtained by PCR amplification of RNA isolated from DLD-1 colon
CC carcinoma cells (ATCC CCL 221). It contains a 7 bp deletion.
CC compared with native SAG, starting at nucleotide 170 of the coding
CC region that codes for a potential haem binding site. The frame
CC shift deletion abolishes the downstream zinc finger in the
CC resulting protein (see AAY06493). The mutation was detected by PCR
CC in SAG RNA, but not in genomic DNA. It suggests a possible role
CC for SAG in human carcinogenesis. SAG genes, and mutant SAG
CC genes, can be used to protect cells from apoptosis induced by redox
CC reagents. Antisense SAG genes can be used to inhibit the growth of
CC tumour cells. The SAG genes can also be used for the recombinant
CC production of the SAG proteins. The SAG proteins can be used to
CC scavenge oxygen radicals in organisms and to promote wound healing.
CC Additionally, the SAG genes or their complements can be used to
CC promote or inhibit the growth of plant cells (all claimed). The SAG
CC protein is also an ideal molecular target in the development of
CC drugs against neurodegenerative disorders, cancers and muscle
CC dystrophy.
XX
XX Sequence 747 BP; 204 A; 155 C; 197 G; 191 T; 0 other;
SQ
Query Match 72.6%; Score 90; DB 20; Length 747;
Best Local Similarity 100.0%; Pred. No. 9.7e-38;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 GATGCTGCTTAGATGTAAGCTGAAACAAACAAGAGACTGTGTGCTGGGGA 94
DB 168 GATGCTGCTTAGATGTAAGCTGAAACAAACAAGAGACTGTGTGCTGGGGA 227
QY 95 GATGTAATCATTCCTCCACAACACTGCTGC 124
DB 228 GATGTAATCATTCCTCCACAACACTGCTGC 257

```

RESULT 22  
AAH25847  
ID AAH25847 standard; DNA; 1152 BP.  
XX  
AC AAH25847;  
XX  
DT 20-AUG-2001 (first entry)  
XX  
DE Human apoptosis associated protein 12 coding sequence.  
XX  
KW Human; apoptosis associated protein 12; cancer;  
KW haemopathy; HIV infection; immunological disease; inflammation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200132863-A1.

```

XX PD 10-MAY-2001.
XX PF 30-OCT-2000; 2000WO-CN00406.
XX PR 29-OCT-1999; 99CN-0119924.
XX (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.
XX PA Mao Y, Xie Y;
XX PI WPI; 2001-335831/35.
XX DR P-PSDB; AAB98975.
XX PT Human apoptosis associated protein 12 and encoded polynucleotide,
XX PT applicable in diagnosis and treatment of malignant tumour, hemopathy,
XX PT HIV infection, immunological diseases and various inflammation -
XX PS Claim 6; Page 22-23; 27pp; Chinese.
XX CC The present invention provides the protein and coding sequences of human
XX CC apoptosis associated (SAG) protein 12. These sequences can be used in
XX CC the diagnosis and treatment of malignant tumours, haemopathy, HIV
XX CC infection, immunological diseases and various types of inflammation. The
XX CC present sequence is the SAG protein 12 coding sequence.
SQ Sequence 1152 BP; 292 A; 265 C; 309 G; 286 T; 0 other;
Query Match 71.8%; Score 89; DB 22; Length 1152;
Best Local Similarity 100.0%; Pred. No. 3.2e-37;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 ATGCCCTGTTAGATGTCAGTGAAGCAACAGAGACTGTGTGCTGGGAG 95
DB 564 ATGCCCTGTTAGATGTCAGTGAAGCAACAGAGACTGTGTGCTGGGAG 623
QY 96 AATGTAATCATTCCTCCACAACTGCTGC 124
DB 624 AATGTAATCATTCCTCCACAACTGCTGC 652
RESULT 23
AAK87320
ID AAK87320 standard; cDNA; 754 BP.
XX AC AAK87320;
XX DT 27-SEP-1999 (first entry)
XX DE Human sensitive to apoptosis (SAG) gene mutant MM4.
XX KW SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX KW neurodegenerative disease; muscular dystrophy; wound healing;
XX OS Homo sapiens.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT CDS 1..342
XX FT /*tag= a
XX FT replace(181,T)
XX FT /*tag= b
XX FT /note= "C61S mutation"
XX PN WO9932514-A2.
XX PD 01-JUL-1999.
XX PF 15-DEC-1998; 98WO-US26705.
XX PR 11-SEP-1998; 98US-0099840.
XX PR 19-DEC-1997; 97US-0068179.

```

```

XX PA (WARN) WARNER LAMBERT CO.
XX PI Sun Y;
XX DR WPI; 1999-430152/36.
XX DR P-PSDB; AAY06498.
XX PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX PT promoting cell growth and protecting cells against apoptosis
XX PS Claim 15; Page 64-65; 84pp; English.
XX CC This is the nucleotide sequence of human sensitive to apoptosis
XX CC (SAG) mutant gene MM4, which codes for a SAG protein (see AAY06498)
XX CC in which the Cys residue at position 61 of the native protein (see
XX CC AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX CC 50 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX CC cDNA. This residue is at zinc-finger finger 1 of SAG. Single and
XX CC double SAG mutants (see AAK87317-31) were made in order to determine
XX CC the role of each cysteine residue of SAG in haem binding and SAG
XX CC oligomerization. These properties were unaffected by the MM4
XX CC mutation. SAG is a novel zinc finger protein that promotes
XX CC cell growth, protects cells from apoptosis, scavenges oxygen
XX CC radicals and can be used for the reversion of a tumour phenotype.
XX CC SAG genes, and mutant SAG genes, can be used to protect cells from
XX CC apoptosis induced by redox reagents. They can also be used for the
XX CC recombinant production of SAG proteins, which are molecular targets
XX CC in the development of drugs against neurodegenerative disorders,
XX CC cancers and muscle dystrophy, and promoting wound healing.
SQ Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
Query Match 66.9%; Score 83; DB 20; Length 754;
Best Local Similarity 100.0%; Pred. No. 4.9e-34;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 42 GTCTTAGATGTCAGTGAAGCAACAGAGACTGTGTGCTGGGAGAAATGA 101
DB 182 GTCTTAGATGTCAGTGAAGCAACAGAGACTGTGTGCTGGGAGAAATGA 241
QY 102 ATCAATCCTTCACAACTGCTGC 124
DB 242 ATCAATCCTTCACAACTGCTGC 264

```

```

RESULT 24
AAK87328
ID AAK87328 standard; cDNA; 754 BP.
XX AC AAK87328;
XX DT 27-SEP-1999 (first entry)
XX DE Human sensitive to apoptosis (SAG) gene mutant MM12.
XX KW SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX KW neurodegenerative disease; muscular dystrophy; wound healing;
XX OS Homo sapiens.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT CDS 1..342
XX FT /*tag= a
XX FT replace(217,T)
XX FT /*tag= b
XX FT /note= "C73S mutation"
XX PN WO9932514-A2.
XX PD 01-JUL-1999.

```



```

FH Key Location/Qualifiers
FT CDS 1..342
FT mutation /*tag= a
FT mutation /replace(181,T)
FT mutation /*tag= b
FT mutation /note= "G61S mutation"
FT mutation /replace(190,T)
FT mutation /*tag= c
FT mutation /note= "C64S mutation"
XX
XX WO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98MO-US26705.
XX
XX 11-SEP-1998; 98US-009840.
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX MPI; 1999-430152/36.
XX P-PSDB; AAY06507.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 79-80; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM13, which codes for a SAG protein (see AAY06507)
XX in which the Cys residues at positions 61 and 64 of the native
XX protein (see AAY06492) are replaced by Ser residues owing to
XX mutations of codons 61 and 64 through site-directed mutagenesis of
XX SAG cDNA. These residues are in zinc ring finger 1 of SAG. Single
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. MM13 showed greatly reduced haem binding (and a
XX decreased brown colour) but oligomerization was unaffected. SAG is
XX a novel zinc finger protein that promotes cell growth, protects cells
XX from apoptosis, scavenges oxygen radicals and can be used for the
XX reversion of a tumour phenotype. SAG genes, and mutant SAG genes,
XX can be used to protect cells from apoptosis induced by redox
XX reagents. They can also be used for the recombinant production of
XX SAG proteins, which are molecular targets in the development of
XX drugs against neurodegenerative disorders, cancers and muscle
XX dystrophy, and promoting wound healing.
XX
XX SQ Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;
XX
XX Query Match 59.7%; Score 74; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-29;
XX Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 51 GTCACGCTGAACAAACAAGAGACTGTGTGTGCTGCGGAGAAATGATTCATTCT 110
XX |||||
XX Db 191 GTCACGCTGAACAAACAAGAGACTGTGTGTGCTGCGGAGAAATGATTCATTCT 250
XX |||||
XX QY 111 TCCACACTGCTGC 124
XX |||||
XX Db 251 TCCACACTGCTGC 264
XX |||||
XX
XX RESULT 27
XX ID AAX41085 standard; cDNA; 224 BP.
XX AC AAX41085;
XX AC
XX AC
XX 18-JUN-1999 (first entry)
XX
XX

```

```

DE Human secreted protein 5' EST SEQ ID NO: 297.
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
XX OS Homo sapiens.
XX
XX WO9906554-A2.
XX
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98MO-IB01238.
XX
XX 01-AUG-1997; 97US-0905134.
XX
XX (GSEST ) GENSEST.
XX
XX Duclert A, Dumas Malne Edwards J, Lacroix B;
XX
XX MPI; 1999-153784/13.
XX P-PSDB; AAY12252.
XX
XX New nucleic acids encoding human secreted proteins - obtained from
XX cDNA libraries prepared from kidney, fetal kidney, dystrophic
XX muscle, muscle and heart tissue
XX
XX Claim 1; Page 441; 622pp; English.
XX
XX AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAY01602 and
XX AAY11994 to AAY12260, respectively. The proteins given represent the
XX signal peptide and an N-terminal fragment of a secreted protein. The
XX nucleic acid sequences can be used for developing products for diagnosis and
XX products. They can also be used to develop products for diagnosis and
XX therapy. The proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used
XX for directing extracellular secretion of a polypeptide or the insertion
XX of a polypeptide into a membrane, or importing a polypeptide into
XX a cell.
XX
XX SQ Sequence 224 BP; 41 A; 60 C; 83 G; 38 T; 2 other;
XX
XX Query Match 34.7%; Score 43; DB 20; Length 224;
XX Best Local Similarity 100.0%; Pred. No. 7e-13;
XX Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 CGATACGTGGCGCCATCTGCAGGGTCCAGGTGATGATGCTGT 43
XX |||||
XX Db 164 CGATACGTGGCGCCATCTGCAGGGTCCAGGTGATGATGATGCTGT 206
XX |||||
XX
XX RESULT 28
XX ID AAX87316 standard; cDNA; 706 BP.
XX AC AAX87316;
XX AC
XX AC
XX 27-SEP-1999 (first entry)
XX
XX Human sensitive to apoptosis (SAG) gene mutant 2.
XX
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX

```

KW neurodegenerative disease; muscular dystrophy; wound healing;  
 XX vulnerable; therapy; mutant; colon carcinoma; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..294  
 FT /tag= a  
 XX MO9932514-A2.  
 XX 01-JUL-1999.  
 XX 15-DEC-1998; 98WO-US26705.  
 XX 11-SEP-1998; 98US-0099840.  
 XX 19-DEC-1997; 97US-0068179.  
 XX (WARN ) WARNER LAMBERT CO.  
 XX Sun Y;  
 XX WPI; 1999-430152/36.  
 XX P-PSDB; AAY06494.  
 PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
 PT promoting cell growth and protecting cells against apoptosis  
 XX Claim 15; Page 55-56; 84pp; English.  
 PS This is the nucleotide sequence of deletion mutant 2 of the novel  
 CC human sensitive to apoptosis gene SAG (see also AAX87314). The cDNA  
 CC was obtained by PCR amplification of RNA isolated from DLD-1 colon  
 CC carcinoma cells (ATCC CCL 221). It contains a 48 bp deletion  
 CC compared with native SAG, starting at nucleotide 177 of the coding  
 CC region that codes for a potential haem binding site. The in-frame  
 CC deletion eliminates 16 amino acids in the encoded protein (see  
 CC AAY06494) but retains the zinc finger motif. The mutation was  
 CC detected by PCR in SAG RNA, but not in genomic DNA. It suggests a  
 CC possible role for SAG in human carcinogenesis. SAG genes, and  
 CC mutant SAG genes, can be used to protect cells from apoptosis  
 CC induced by redox reagents. Antisense SAG genes can be used to  
 CC inhibit the growth of tumour cells. The SAG genes can also be used  
 CC for the recombinant production of the SAG proteins. The SAG  
 CC proteins can be used to scavenge oxygen radicals in organisms and  
 CC to promote wound healing. Additionally, the SAG genes or their  
 CC complements can be used to promote or inhibit the growth of plant  
 CC cells (all claimed). The SAG protein is also an ideal molecular  
 CC target in the development of drugs against neurodegenerative  
 CC disorders, cancers and muscle dystrophy.  
 XX Sequence 706 BP; 189 A; 147 C; 189 G; 181 T; 0 other;  
 SQ  
 Query Match 33.9%; Score 42; DB 20; Length 706;  
 Best Local Similarity 100.0%; Pred. No. 2,3e-12;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 GTGGCTGGGGAGAAATGTAATCATCTTCCTCCACAACTGCTGC 124  
 DB 175 GTGGCTGGGGAGAAATGTAATCATCTTCCTCCACAACTGCTGC 216  
 RESULT 29  
 AAC77493  
 ID AAC77493 standard; cDNA; 441 BP.  
 XX AAC77493;  
 AC AAC77493;  
 XX 08-FEB-2001 (first entry)  
 DT Human ORFX ORF3048 polynucleotide sequence SEQ ID NO:6095.  
 DE Human; open reading frame; ORFX; detection; cyostatic; hepatotropic;

KW vulnerable; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200058473-A2.  
 XX 05-OCT-2000.  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX 31-MAR-1999; 99US-0127607.  
 XX 02-APR-1999; 99US-0127636.  
 XX 05-APR-1999; 99US-0127728.  
 XX 30-MAR-2000; 2000US-0540763.  
 XX (CURA-) CURAGEN CORP.  
 XX Shinkens RA, Leach M;  
 XX WPI; 2000-602362/57.  
 XX P-PSDB; AAB43284.  
 DR Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX Claim 5; Page 5279; 5507pp; English.  
 PS  
 XX AACT74446 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cyostatic; hepatotropic; vulnerable;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihypoid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX Sequence 441 BP; 98 A; 113 C; 139 G; 90 T; 1 other;  
 SQ  
 Query Match 33.1%; Score 41; DB 21; Length 441;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-12;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 GTGGCTGGGGAGAAATGTAATCATCTTCCTCCACAACTGCTG 123  
 DB 208 GTGGCTGGGGAGAAATGTAATCATCTTCCTCCACAACTGCTG 248

RESULT 30  
ABN40537  
ID ABN40537 standard; DNA; 60 BP.  
XX  
AC ABN40537;  
XX  
XX 15-JUL-2002 (first entry)  
XX  
DE Human spliced transcript detection oligonucleotide SEQ ID NO:13285.  
XX  
KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200210449-A2.  
XX  
PD 07-FEB-2002.  
XX  
XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
XX  
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX  
XX MPI; 2002-257383/30.  
XX  
PT New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes  
XX  
XX Example 1; SEQ ID 13285; 47bp; English.  
XX  
XX The present invention describes oligonucleotide libraries for detecting  
XX messenger RNAs that populate a (sub-)transcriptome, where the  
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
XX transcription units that populate a genome. The library comprises  
XX several oligonucleotides, each capable of hybridizing selectively to a  
XX set of messenger RNAs transcribed from a given transcription unit of  
XX the genome, which encodes one or more messenger RNA splice variants.  
XX The oligonucleotide libraries are useful for detecting mRNAs from a  
XX biological sample, in expression profiling studies, in qualitatively or  
XX quantitatively characterizing the corresponding transcriptome, and in  
XX detecting RNA transcripts and splice variants of human or animal  
XX transcriptomes. The libraries may also be used as specialised mini  
XX libraries to detect transcripts of a sub-transcriptome under a  
XX particular biological or pathological state, and so allowing the  
XX detection of tissue- and pathology-specific genes such as those genes  
XX only expressed in specific tissue under a specific pathological  
XX condition; to detect developmental specific genes; and to detect RNA  
XX transcripts and splice variants of a transcriptome of a patient suffering  
XX from a particular disorder. ABN27253 to ABN55589 represent  
XX oligonucleotide sequences from rats, humans and mice, which are used in  
XX the exemplification of the present invention.  
XX N.B. The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 60 BP; 10 A; 13 C; 25 G; 12 T; 0 other;  
XX  
Query Match 28.2%; Score 35; DB 24; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGGCCATCTGCGAGGTCAGGTGATGG 35  
DB 26 CGATACGTGGCCATCTGCGAGGTCAGGTGATGG 60

RESULT 31  
ABZ11414  
ID ABZ11414 standard; CDNA; 439 BP.  
XX  
XX  
AC ABZ11414;  
XX  
XX 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 296.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; neurotropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200270539-A2.  
XX  
XX 12-SEP-2002.  
XX  
XX 05-MAR-2002; 2002WO-US05095.  
XX  
XX 05-MAR-2001; 2001US-0799451.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YF, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
XX Weinman T, Wang J, Wang D, Drmanac RT;  
XX  
XX MPI; 2002-759812/82.  
XX  
XX P-PSDB; ABP69197.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
XX sequence tags (ESTs), useful for treating cell-proliferative,  
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
XX platelet or coagulation disorders  
XX  
XX Claim 1; SEQ ID NO 296; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
XX nucleotide sequence selected from any of 948 sequences  
XX (ABZ1119-ABZ12066) or their mature protein coding portion, active domain  
XX coding protein or complementary sequences. The polynucleotides are useful  
XX for identifying expressed genes or for physical mapping of human genome.  
XX The encoded polypeptides (ABP68902-ABP69843) are useful as molecular  
XX weight markers, as a food supplement, for generating antibodies, in  
XX medical imaging, screening and diagnostic assays and for treating  
XX cell-proliferative disorders (cancer), neurodegenerative diseases  
XX (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
XX sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
XX disorders, platelet or coagulation disorders, wound, burns, infection,  
XX ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
XX parasitic), arthritis, etc.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 439 BP; 69 A; 142 C; 162 G; 66 T; 0 other;  
XX

QY 1 CGATACGTGGCCATCTGCGAGGTCAGGTGATGG 35  
DB 221 CGATACGTGGCCATCTGCGAGGTCAGGTGATGG 255

RESULT 32  
AA87313  
ID AA87313 standard; cDNA; 1140 BP.  
AC  
XX  
AC AA87313;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Mouse sensitive to apoptosis (SAG) gene.  
XX  
KM SAG gene; sensitive to apoptosis; mouse; cancer; tumour;  
KM neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerable; therapy; ds.  
XX  
OS Mus musculus.  
FH  
FH Key Location/Qualifiers  
FT CDS 17..358  
FT /\*tag= a  
XX  
XX MO9932514-A2.  
XX  
XX 01-JUL-1999.  
XX  
XX 15-DEC-1998; 98MO-US26705.  
XX  
XX 11-SEP-1998; 98US-0099840.  
XX 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN) WARNER LAMBERT CO.  
XX  
XX  
XX Sun Y;  
XX  
XX WPI; 1999-430152/36.  
XX P-PSDB; AA06491.  
XX  
XX  
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
XX  
XX Claim 1; Page 48-49; 84pp; English.  
XX  
XX  
XX This is the nucleotide sequence of mouse sensitive to apoptosis  
CC (SAG) cDNA, which codes for a novel redox-sensitive, hem-binding  
CC protein (see AA06491) with a zinc RING finger domain that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversal of a tumour phenotype.  
CC The cDNA was isolated using differential display to identify genes  
CC associated with 1,10-phenanthroline (OP)-induced apoptosis in  
CC murine tumour lines, and use of an isolated OP-inducible clone to  
CC screen a mouse lung cDNA library. The mouse SAG cDNA was used to  
CC identify human SAG (see AA87314). SAG is highly conserved among  
CC species. Disruption in yeast was shown to be lethal. SAG deletion  
CC mutants (see AA87315-16) have been identified in human cancer lines,  
CC suggesting a role in carcinogenesis. SAG genes, and mutant SAG  
CC genes, can be used to protect cells from apoptosis induced by redox  
CC reagents. Antisense SAG genes can be used to inhibit the growth of  
CC tumour cells. The SAG genes can also be used for the recombinant  
CC production of the SAG proteins. The SAG proteins can be used to  
CC scavenge oxygen radicals in organisms and to promote wound healing.  
CC Additionally, the SAG genes or their complements can be used to  
CC promote or inhibit the growth of plant cells (all claimed). The SAG  
CC protein is also an ideal molecular target in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy.  
XX  
XX  
SQ Sequence 1140 BP; 302 A; 224 C; 287 G; 327 T; 0 other;

Query Match 25.8%; Score 32; DB 20; Length 1140;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCATCTGCAGGGTCCAGGTGATGATGCTG 42

Db  
167 GCCATCTGCAGGGTCCAGGTGATGATGCTG 198  
|||||  
RESULT 33  
AAL16209  
ID AAL16209 standard; cDNA; 264 BP.  
XX  
XX  
XX AAL16209;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human breast cancer expressed polynucleotide 8666.  
XX  
XX Human; breast cancer; cell marker; cyrostatic; ss.  
XX  
XX Homo sapiens.  
XX  
XX MO200151628-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 10-JAN-2001; 2001WO-US00798.  
XX  
XX 14-JAN-2000; 2000US-0176077.  
XX 14-MAR-2000; 2000US-0189167.  
XX 24-MAR-2000; 2000US-0192089.  
XX 29-MAR-2000; 2000US-0193480.  
XX 15-MAY-2000; 2000US-0205230.  
XX 09-JUN-2000; 2000US-0211315.  
XX 25-JUL-2000; 2000US-0220534.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
XX WPI; 2001-451856/48.  
XX  
XX  
XX New peptide useful as a marker for the diagnosis of breast cancer  
XX  
XX  
XX Claim 1; Page 1568; 3695pp; English.  
XX  
XX  
XX The invention relates to human breast cancer expressed polynucleotides  
CC (AAL07544-AAL26783) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterising treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cyrostatic  
CC activity.  
XX  
XX  
SQ Sequence 264 BP; 64 A; 59 C; 74 G; 62 T; 5 other;

Query Match 19.4%; Score 24; DB 22; Length 264;  
Best Local Similarity 100.0%; Pred. No. 0.0078;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 AATCATCTCTTCCACACTGCTGC 124  
|||||  
Db 54 AATCATCTCTTCCACACTGCTGC 77  
|||||

RESULT 34  
AAL25052  
ID AAL25052 standard; cDNA; 596 BP.  
XX  
XX  
XX AAL25052;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human breast cancer expressed polynucleotide 17509.  
XX

```
KW Human; breast cancer; cell marker; cyrostatic; ss.
XX
XX Homo sapiens.
OS
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX
XX 14-MAR-2000; 2000US-0189167.
XX
XX 24-MAR-2000; 2000US-0192099.
XX
XX 29-MAR-2000; 2000US-0193480.
XX
XX 15-MAY-2000; 2000US-0205230.
XX
XX 09-JUN-2000; 2000US-0211315.
XX
XX 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX Claim 1; Page 3235; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AA107544-AA126789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cyrostatic
XX activity.
XX
XX Sequence 596 BP; 166 A; 124 C; 144 G; 162 T; 0 other;
XX
XX Query Match 19.4%; Score 24; DB 22; Length 596;
XX Best Local Similarity 100.0%; Pred. No. 0.0077;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 101 AATCATTCCTTCACAACTGCTGC 124
XX |||||||||||||||||||
XX Db 11 AATCATTCCTTCACAACTGCTGC 34
XX
XX RESULT 35
XX AAL22239
XX ID AAL22239 standard; cDNA; 630 BP.
XX
XX AAL22239;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 14696.
XX
XX Human; breast cancer; cell marker; cyrostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX
XX 14-MAR-2000; 2000US-0189167.
XX
XX 24-MAR-2000; 2000US-0192099.
XX
XX 29-MAR-2000; 2000US-0193480.
XX
XX 15-MAY-2000; 2000US-0205230.
XX
XX 09-JUN-2000; 2000US-0211315.
XX
XX 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX Claim 1; Page 3235; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AA107544-AA126789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cyrostatic
XX activity.
XX
XX Sequence 596 BP; 166 A; 124 C; 144 G; 162 T; 0 other;
XX
XX Query Match 19.4%; Score 24; DB 22; Length 596;
XX Best Local Similarity 100.0%; Pred. No. 0.0077;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX Claim 1; Page 2649-2650; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AA107544-AA126789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cyrostatic
XX activity.
XX
XX Sequence 630 BP; 213 A; 125 C; 121 G; 171 T; 0 other;
XX
XX Query Match 19.4%; Score 24; DB 22; Length 630;
XX Best Local Similarity 100.0%; Pred. No. 0.0077;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 101 AATCATTCCTTCACAACTGCTGC 124
XX |||||||||||||||||||
XX Db 43 AATCATTCCTTCACAACTGCTGC 66
XX
XX RESULT 36
XX AAA70390/C
XX ID AAA70390 standard; DNA; 45 BP.
XX
XX AAA70390;
XX
XX 02-FEB-2001 (first entry)
XX
XX Site directed mutagenesis oligonucleotide # 1 for placental bikunin.
XX
XX Mucociliary dysfunction; mucus; sputum; human;
XX chronic obstructive lung disease; chronic bronchitis; CB; Bronchiectasis;
XX BE; asthma; cystic fibrosis; CF; bacterial infection; placental bikunin;
XX Kunitz-type serine protease inhibitor; chronic sinusitis; glue ear; ss.
XX
XX Homo sapiens.
XX
XX WO200037099-A2.
XX
XX 29-JUN-2000.
XX
XX 22-DEC-1999; 99WO-GB04381.
XX
XX 22-DEC-1998; 98US-0218913.
XX
XX 17-NOV-1999; 99US-0441966.
XX
XX (FARB ) BAYER AG.
XX
XX Hall R, Poll CT, Newton BB, Taylor WJA;
XX
XX WPI; 2000-452127/39.
XX
XX Stimulating mucociliary clearance rate of mucus and sputum in lung
XX airways for treating lung diseases such as cystic fibrosis and
XX bronchitis involves administering a Kunitz-type serine protease
XX inhibitor -
XX
XX Example 9; Page 59; 173pp; English.
```



XX Mucociliary dysfunction is the inability of ciliated epithelium to clear  
 CC mucus and sputum in lung airways. Mucociliary dysfunction is a serious  
 CC complication of chronic obstructive lung diseases such as Chronic  
 CC Bronchitis (CB), Bronchiectasis (BE), asthma and Cystic Fibrosis (CF).  
 CC In addition, patients suffering from mucociliary dysfunction are  
 CC susceptible to secondary bacterial infections. A sequence for human  
 CC placental bikunin has been isolated (see AAB14190 for the protein  
 CC sequence). Placental bikunin is a Kunitz-type serine protease inhibitor  
 CC and protein, which can stimulate the rate of mucociliary clearance of mucus  
 CC and sputum in lung airways. Therefore, placental bikunin protein may be  
 CC used for treating lung diseases such as CF, CB, BE, and chronic  
 CC sinusitis and glue ear which are caused by retention and accumulation of  
 CC mucus. The present sequence is an oligonucleotide used in site-directed  
 CC mutagenesis of the coding sequence of the human placental bikunin protein  
 CC sequence of AAB14190.

XX  
 XX  
 SQ Sequence 45 BP; 4 A; 15 C; 18 G; 8 T; 0 other;

Query Match 15.3%; Score 19; DB 21; Length 45;  
 Best Local Similarity 100.0%; Pred. No. 3.5;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ACGTGGCCCATCTGCAGG 23  
 |||||  
 38 ACGTGGCCCATCTGCAGG 20

Db

RESULT 37  
 AAX87337  
 ID AAX87337 standard; DNA; 18 BP.  
 XX  
 AC AAX87337;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX  
 DE Human sensitive to apoptosis gene (SAG) primer hSAG.M1.  
 XX  
 XX SAG gene: sensitive to apoptosis; human; cancer; tumour;  
 KM neurodegenerative disease; muscular dystrophy; wound healing;  
 KM vulnerability; therapy; PCR; primer; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 OS  
 PN MO932514-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PF 15-DEC-1998; 98MO-US26705.  
 XX  
 PR 11-SEP-1998; 98US-0099840.  
 PR 19-DEC-1997; 97US-0068179.  
 XX  
 PA (WARN) WARNER LAMBERT CO.  
 XX  
 PI Sun Y;  
 PI  
 DR WPI; 1999-430152/36.  
 XX  
 XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
 PT promoting cell growth and protecting cells against apoptosis  
 PT  
 XX Example 13; Page 26; 84pp; English.  
 PS  
 CC Primer hSAG.M1 corresponds to nucleotides 151-171 of the human  
 CC sensitive to apoptosis gene (SAG) cDNA clone provided in AAX87314.  
 CC It was used with primer SAGT-02-1 (see AAX87338) in the RT-PCR  
 CC amplification of RNA isolated from 20 human tumour lines and  
 CC transformed lines originating from lung, brain, kidney, prostate,  
 CC testis, nasopharynx, bone, cervix and foreskin. 2 SAG deletion  
 CC mutants (see AAX87315-16) were detected in cancer cell lines  
 CC originating from colon and testis, suggesting a possible role for

CC SAG in human carcinogenesis. SAG (see also AAY06492) is a  
 CC redox-sensitive, haem-binding protein that promotes cell growth. It  
 CC protects cells from apoptosis, and scavenges oxygen radicals. It  
 CC can be used to reverse a tumour phenotype.  
 XX  
 SQ Sequence 18 BP; 3 A; 6 C; 6 G; 3 T; 0 other;

Query Match 14.5%; Score 18; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GCCATCTGCAGGTCAG 28  
 |||||  
 1 GCCATCTGCAGGTCAG 18

Db

RESULT 38  
 AAC29980/c  
 ID AAC29980 standard; cDNA; 349 BP.  
 XX  
 AC AAC29980;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 34055.  
 XX  
 DE Human 5' EST, expressed sequence tag; secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI  
 DR WPI; 2000-500381/45.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 34055; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SQ Sequence 349 BP; 101 A; 65 C; 67 G; 116 T; 0 other;

Query Match 14.5%; Score 18; DB 21; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 GGGAGATGATATCATTC 108

Db 85 GGGAGATGTATCATTC 68

RESULT 39  
ABV87139/c  
ID ABV87139 standard; cDNA; 541 BP.

XX ABV87139;

DT 13-DEC-2002 (first entry)

XX Human colon cancer related cDNA SEQ ID NO 450.

DE Human colon cancer; cancer; cytostatic; tumour; gene therapy; vaccine; gene;

XX ss.

XX Homo sapiens.

XX MO200258534-A2.

XX 01-AUG-2002.

XX 19-NOV-2001; 2001WO-US43704.

XX 20-NOV-2000; 2000US-252222P.

XX 06-FEB-2001; 2001US-267011P.

XX 28-MAR-2001; 2001US-279670P.

XX 10-JUL-2001; 2001US-304037P.

XX (CORI-) CORIXA CORP.

XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;

XX WPI; 2002-608400/65.

XX New isolated tumor colon polynucleotide and polypeptide, useful for the

XX diagnosis, prevention and/or treatment of cancer, in particular colon

XX cancer -

XX Claim 1; SEQ ID NO 450; 266pp + Sequence Listing; English.

XX The invention relates to a human colon tumour expressed polynucleotide

XX CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of

XX CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)

XX CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)

XX CC sequences that hybridize to (i); under moderately stringent conditions;

XX CC (v) sequences having at least 75% or 90% identity to (i); or (vi)

XX CC degenerate variants of (i). The compositions and methods of the present

XX CC invention are useful for the diagnosis, prevention and/or treatment of

XX CC cancer, particularly colon cancer. (I) can be used in gene therapy and

XX CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 541 BP; 173 A; 99 C; 105 G; 164 T; 0 other;

XX Query Match 14.5%; Score 18; DB 24; Length 541;

XX Best Local Similarity 100.0%; Pred. No. 11;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108

DB 154 GGGAGATGTATCATTC 137

RESULT 40

AAQ70827/c

XX AAQ70827 standard; cDNA; 2405 BP.

XX AAQ70827;

DT 25-MAR-2003 (updated)

DT 17-MAR-1995 (first entry)

XX Protein kinase (CKIgamma1Hu) coding sequence.

XX Protein kinase; immunogen; antibody; protein-tyrosine kinase;

XX protein-serine/threonine kinase; recombination; repair; screening;

XX detection; casein kinase; ss.

XX Homo sapiens.

XX Key

XX CDS

XX Location/Qualifiers

XX 67..1200

XX /tag= a

XX /product= Protein kinase.

XX MO9417189-A2.

XX 04-AUG-1994.

XX 21-JAN-1994; 94WO-US00795.

XX 21-JAN-1993; 93US-0008001.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Hoeftstra MF;

XX WPI; 1994-264102/32.

XX P-PSDB; AAR56529.

XX Wild-type and mutant protein kinase genes and encoded

XX polypeptide(s) - Useful in screening for compositions which may

XX effect DNA double strand break repair activity

XX Claim 10; Page 97-99; 121pp; English.

XX This sequence encodes a protein kinase which is an HRR25-like

XX protein. Host cells stably transformed with the protein kinase

XX encoding DNA may be used for the expression of the protein kinase

XX such that the expressed protein is "displayed" on the host cell

XX surface. The cells may then be used as immunogen for the production

XX of antibodies. The host cells may also be used for the large scale

XX production of the protein kinase, the expressed protein being either

XX isolated from the cell surface or from the culture medium.

XX Recombinant HRR25 like proteins display a number of properties which

XX are unique among the eukaryotic protein kinases e.g. the HRR25

XX CC protein possesses both protein-tyrosine kinase and

XX CC protein-serine/threonine kinase activities. Also, HRR25 operates to

XX CC promote repair of DNA strand breaks at a specific nucleotide sequence

XX CC and is the only protein kinase known to have such

XX CC recombination/repair promoting activity. Recombinant HRR25-like

XX CC proteins and host cells expressing them are useful in screening

XX CC methods designed to examine the effects of various compositions on

XX CC DNA break repair and protein kinase activities of the protein. The

XX CC HRR25-like proteins are casein kinase class I protein kinases.

XX CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 2405 BP; 849 A; 392 C; 486 G; 678 T; 0 other;

XX Query Match 14.5%; Score 18; DB 15; Length 2405;

XX Best Local Similarity 100.0%; Pred. No. 11;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108

DB 2241 GGGAGATGTATCATTC 2224

RESULT 41

AAQ92962/c

XX AAQ92962 standard; DNA; 2405 BP.

XX AAQ92962;

```

AC AA092962;
XX
XX 25-JAN-1996 (first entry)
XX
XX Human HRR25-like casein-kinase-I CKI-gamma-1Hu gene.
DE
XX Casein-kinase-I; HRR25; protein-tyrosine-kinase; CKI-gamma-1Hu;
KW protein-serine/threonine-kinase; enzyme; DNA repair;
XX DNA recombination; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 67..1197
XX FT /*tag= a
XX
XX WO9519993-A1.
XX
XX 27-JUL-1995.
XX
XX 23-JAN-1995; 95WO-US00955.
XX
XX 21-JAN-1994; 94US-0185359.
XX
XX (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
XX Hoekstra MF;
XX
XX WPI; 1995-269419/35.
XX
XX P-PSDB; AAR76625.
XX
XX New monoclonal antibodies against human casein Kinase class I
XX PT enzymes - useful for purification and determination of these enzymes and to
XX PT modulate their receptor-ligand binding, also new hybridomas
XX
XX Disclosure; Page 100; 125pp; English.
XX
XX This human casein-kinase-I (CKI-gamma-1Hu) gene which is a homolog
XX CC of a DNA recombination and repair gene called HRR25 (Ho and/or
XX CC radiation repair) possesses both protein-tyrosine-kinase and protein
XX CC serine/threonine-kinase activities. HRR25 operates to promote repair
XX CC of DNA strand breaks at a specific nucleotide sequence and is the
XX CC only protein-kinase known to have such recombination/repair
XX CC promoting activity. DNA encoding HRR25 can be used to isolate and
XX CC identify related sequences, while recombinant HRR25, or cells
XX CC expressing it, can be used to screen compounds for their effects
XX CC on DNA repair and kinase activities.
XX
XX Sequence 2405 BP; 849 A; 392 C; 486 G; 678 T; 0 other;
SQ
Query Match 14.5%; Score 18; DB 16; Length 2405;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 91 GGGAGATGTATCATTC 108
DB 2241 GGGAGATGTATCATTC 2224

```

```

XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX PI WPI; 2001-235357/24.
XX
XX P-PSDB; AAG74302.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PT
XX Claim 1; Page 2758; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patient's own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated P,
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAB77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2630 BP; 882 A; 416 C; 435 G; 891 T; 6 other;
SQ
Query Match 14.5%; Score 18; DB 22; Length 2630;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 91 GGGAGATGTATCATTC 108
DB 1143 GGGAGATGTATCATTC 1126

```

```

RESULT 42
AAH3733/c
ID AAH3733 standard; cDNA; 2630 BP.
XX
XX AAH3733;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:789.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma; ss.
XX
OS Homo sapiens.

```

```

RESULT 43
AAS84630/c
ID AAS84630 standard; cDNA; 239 BP.
XX
XX AAS84630;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #20434.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.

```

XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Dmanac RT, Liu C, Tang YT;  
PI MPI; 2001-639362/73.  
DR P-PSDB; ABG20443.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PS Claim 1, SEQ ID No 20434; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. A564197-A564564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 239 BP; 47 A; 87 C; 58 G; 47 T; 0 other;  
Query Match 13.7%; Score 17; DB 23; Length 239;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 CCATCTGCAGGGTCCAG 28  
Db 139 CCATCTGCAGGGTCCAG 123  
RESULT 44  
ABA46329  
ID ABA46329 standard; DNA; 355 BP.  
XX  
XX ABA46329;  
AC 01-FEB-2002 (first entry)  
XX  
XX  
DT 01-FEB-2002 (first entry)  
XX  
XX Human breast cell single exon nucleic acid probe #5024.  
DE Human breast cell single exon nucleic acid probe #5024.  
XX  
XX Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157271-A2.  
FN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00662.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0633366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI MPI; 2001-496933/54.  
XX  
XX  
XX  
XX  
XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes  
PS Claim 1, SEQ ID NO 5024; 327pp + sequence listing; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
Query Match 13.7%; Score 17; DB 22; Length 355;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 GGTCCAGGTGATGATG 38  
Db 281 GGTCCAGGTGATGATG 297  
RESULT 45  
ABA51431  
ID ABA51431 standard; DNA; 355 BP.  
XX  
XX ABA51431;  
AC 01-FEB-2002 (first entry)  
XX  
XX  
DT 01-FEB-2002 (first entry)  
XX  
XX Human breast cell single exon nucleic acid probe #10126.  
DE Human breast cell single exon nucleic acid probe #10126.  
XX  
XX Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157271-A2.  
FN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00662.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR

PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 PS Claim 4; SEQ ID NO 10126; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
 SQ  
 XX  
 Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 22 GGTCGAGTGATGATG 38  
 |||||  
 DB 281 GGTCGAGTGATGATG 297  
 RESULT 46  
 ABA56889  
 ID ABA56889 standard; DNA; 355 BP.  
 XX  
 AC ABA56889;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #5194.  
 XX  
 DE Human foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 OS MO200157277-A2.  
 PN  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 1; SEQ ID NO 5194; 639bp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
 SQ  
 XX  
 Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 22 GGTCGAGTGATGATG 38  
 |||||  
 DB 281 GGTCGAGTGATGATG 297  
 RESULT 47  
 ABA69454  
 ID ABA69454 standard; DNA; 355 BP.  
 XX  
 AC ABA69454;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #17759.  
 XX  
 DE Human foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 OS MO200157277-A2.  
 PN  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 4; SEQ ID NO 17759; 639bp + sequence listing; English.  
 XX

CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from [ftp:wipo.int/pub/published\\_pct\\_sequences](ftp:wipo.int/pub/published_pct_sequences).  
 CC  
 SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;

Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 GGTCAGGTGATGATG 38  
 DB 281 GGTCAGGTGATGATG 297

RESULT 48  
 ABA26508  
 ID ABA26508 standard; DNA; 355 BP.  
 XX  
 AC ABA26508;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Probe #4974 for gene expression analysis in human heart cell sample.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 1; SEQ ID No 4974; 530bp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at [ftp:wipo.int/pub/published\\_pct\\_sequences](ftp:wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;

Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 GGTCAGGTGATGATG 38  
 DB 281 GGTCAGGTGATGATG 297

RESULT 49  
 ABA36383  
 ID ABA36383 standard; DNA; 355 BP.  
 XX  
 AC ABA36383;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX

DE Probe #14849 for gene expression analysis in human heart cell sample.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00666.  
 XX

PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -

PS Claim 4; SEQ ID No 14849; 530bp; English.

CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp:wipo.int/pub/published\\_pct\\_sequences](ftp:wipo.int/pub/published_pct_sequences).  
 XX

Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 GGTCCAGTGATGATG 38  
 |||||  
 Db 281 GGTCCAGTGATGATG 297

## RESULT 50

AAK04990  
 ID AAK04990 standard; DNA; 355 BP.  
 XX

AC AAK04990;  
 XX

DT 05-NOV-2001 (first entry)  
 XX

DE Human brain expressed single exon probe SEQ ID NO: 4981.  
 XX

KM Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer; ss.  
 XX

OS Homo sapiens.  
 XX

PN MO200157275-A2.  
 XX

PD 09-AUG-2001.  
 XX

PF 30-JAN-2001; 2001MO-US00667.  
 XX

PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX

DR WPI; 2001-483446/52.  
 XX

PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX

PS Example 4; SEQ ID NO: 4981; 650bp + Sequence Listing; English.  
 XX

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system,  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX

SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;

Query Match 13.7%; Score 17; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred.No. 39;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 GGTCCAGTGATGATG 38  
 |||||  
 Db 281 GGTCCAGTGATGATG 297

Search completed: November 7, 2003, 11:58:14  
 Job time : 127.47 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 05:52:38 / Search time 33.5936 Seconds  
(without alignments)  
1629.223 Million cell updates/sec

Title: US-09-509-779-3\_COPY\_141\_264

Perfect score: 124  
Sequence: 1 CGATACGTGCGCCATCTGCA.....ATTCTTCACACACTGCTGC 124

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfills1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	15.3	45	4	US-09-144-428-61
C 2	18	14.5	2405	1	US-08-454-097-30
C 3	18	14.5	2405	3	US-08-185-359-30
C 4	17	13.7	363	4	US-09-328-352-1220
C 5	17	13.7	3494	3	US-09-139-802-200
C 6	17	13.7	3494	4	US-09-659-786-200
C 7	17	13.7	8654	4	US-08-961-527-98
C 8	16	12.9	266	4	US-09-313-294A-3641
C 9	16	12.9	458	4	US-09-149-476-239
C 10	16	12.9	475	4	US-09-149-476-76
C 11	16	12.9	1059	3	US-08-163-919A-1
C 12	16	12.9	1059	5	PCT-US94-14073-1
C 13	16	12.9	5769	1	US-08-652-971-1
C 14	16	12.9	5769	2	US-08-991-258A-1
C 15	16	12.9	5769	3	US-08-769-399-1
C 16	16	12.9	5769	3	US-08-991-953A-1
C 17	15	12.1	196	3	US-08-663-082-1
C 18	15	12.1	275	4	US-09-313-294A-2588
C 19	15	12.1	819	1	US-08-309-182B-2
C 20	15	12.1	942	4	US-09-252-991A-11847
C 21	15	12.1	978	4	US-09-252-991A-11595
C 22	15	12.1	1137	4	US-09-252-991A-5529
C 23	15	12.1	1196	1	US-08-290-448A-71
C 24	15	12.1	1196	1	US-08-290-448A-71
C 25	15	12.1	1196	1	US-08-175-069A-71
C 26	15	12.1	1196	4	US-08-461-939B-71
C 27	15	12.1	1196	4	US-08-464-000-71

28	15	12.1	1292	4	US-09-205-258-189	Sequence 189, App
29	15	12.1	1314	3	US-09-025-059-2	Sequence 2, Appl
30	15	12.1	1368	1	US-08-290-448A-79	Sequence 79, Appl
31	15	12.1	1368	1	US-08-290-448A-79	Sequence 79, Appl
32	15	12.1	1368	1	US-08-175-069A-79	Sequence 79, Appl
33	15	12.1	1368	4	US-08-461-939B-79	Sequence 79, Appl
34	15	12.1	1368	4	US-08-464-000-79	Sequence 79, Appl
35	15	12.1	1426	3	US-09-121-425-3	Sequence 3, Appl
36	15	12.1	1426	4	US-08-634-493A-3	Sequence 3, Appl
37	15	12.1	1506	1	US-08-149-105-4	Sequence 4, Appl
C 38	15	12.1	1506	1	US-08-317-847-4	Sequence 4, Appl
C 39	15	12.1	1545	4	US-09-252-991A-5553	Sequence 5553, Ap
C 40	15	12.1	1605	4	US-09-252-991A-11780	Sequence 11780, A
C 41	15	12.1	1794	3	PCT-US94-10080-9	Sequence 9, Appl
C 42	15	12.1	1794	5	PCT-US94-10080-9	Sequence 9, Appl
C 43	15	12.1	1914	4	US-09-252-991A-5670	Sequence 5670, Ap
C 44	15	12.1	2308	3	US-09-382-256-9	Sequence 9, Appl
C 45	15	12.1	2308	3	US-09-395-115-9	Sequence 9, Appl

#### ALIGNMENTS

RESULT 1  
US-09-144-428-61/c  
Sequence 61, Application US/09144428  
Patent No. 6583108  
GENERAL INFORMATION:  
APPLICANT: BAYER CORPORATION, The  
APPLICANT: TAMBURINI, Paul P  
APPLICANT: DAVIS, Gary  
APPLICANT: DELARIA, Katherine A  
APPLICANT: MARJOR, Christopher W  
APPLICANT: MULLER, Daniel K  
TITLE OF INVENTION: HUMAN BIKUNIN  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 S. Wacker Drive Suite 3200  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,428  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/03894  
FILING DATE: 10-MAR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,106  
FILING DATE: 11-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/019,793  
FILING DATE: 14-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/725,251  
FILING DATE: 04-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: CHAO, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 96,223-11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 913-0001  
TELEFAX: (312) 913-0002  
INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: not relevant  
MOLECULE TYPE: DNA (genomic)  
US-09-144-428-61

Query Match 15.3%; Score 19; DB 4; Length 45;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 AGTGGCCATCTGCAGG 23  
|||||  
DB 38 AGTGGCCATCTGCAGG 20

## RESULT 2

US-08-454-097-30/C  
Sequence 30, Application US/08454097  
Patent No. 5686412  
GENERAL INFORMATION:  
APPLICANT: Hoechst, Merl F.  
TITLE OF INVENTION: Protein Kinases  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 233 South Wacker Drive, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,097  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/185,359  
FILING DATE: 21-JAN-1994  
APPLICATION NUMBER: US 08/008,001  
FILING DATE: 21-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728,783  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5686412and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31853  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2405 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 67..1197  
US-08-454-097-30

Query Match 14.5%; Score 18; DB 1; Length 2405;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108  
|||||  
DB 2241 GGGAGATGTATCATTC 2224

## RESULT 3

US-08-185-359-30/C  
Sequence 30, Application US/08185359  
Patent No. 6060296  
GENERAL INFORMATION:  
APPLICANT: Hoechst, Merl F.  
TITLE OF INVENTION: Protein Kinases  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 233 South Wacker Drive, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,359  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/008,001  
FILING DATE: 21-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728,783  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6060296and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31853  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2405 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 67..1197  
US-08-185-359-30

Query Match 14.5%; Score 18; DB 3; Length 2405;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGAGATGTATCATTC 108  
|||||  
DB 2241 GGGAGATGTATCATTC 2224

## RESULT 4

US-09-328-352-1220  
Sequence 1220, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Bretton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA

;/ CURRENT APPLICATION NUMBER: US/09/328,352  
;/ CURRENT FILING DATE: 1999-06-04  
;/ NUMBER OF SEQ ID NOS: 8252  
;/ SEQ ID NO 1220  
;/ LENGTH: 363  
;/ TYPE: DNA  
;/ ORGANISM: Acinetobacter baumannii  
US-09-328-1220

Query Match 13.7%; Score 17; DB 4; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 TGTGTGTGTGTGGG 93  
DB 213 TGTGTGTGTGTGGG 229

RESULT 5  
US-09-139-802-200/c  
;/ Sequence 200, Application US/09139802  
;/ Patent No. 6180084  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Ruoslahti, Erkki  
;/ APPLICANT: Pasqualini, Renata  
;/ TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
;/ TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
;/ TITLE OF INVENTION: Same  
;/ FILE REFERENCE: P-LU 3203  
;/ CURRENT APPLICATION NUMBER: US/09/139,802  
;/ CURRENT FILING DATE: 1998-08-25  
;/ EARLIER APPLICATION NUMBER: 08/926,914  
;/ EARLIER FILING DATE: 1997-09-10  
;/ EARLIER APPLICATION NUMBER: 08/710,067  
;/ EARLIER FILING DATE: 1996-09-10  
;/ NUMBER OF SEQ ID NOS: 226  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 200  
;/ LENGTH: 3494  
;/ TYPE: DNA  
;/ ORGANISM: Homo sapiens  
;/ FEATURE:  
;/ NAME/KEY: CDS  
;/ LOCATION: (121)..(3024)  
US-09-139-802-200

Query Match 13.7%; Score 17; DB 3; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCGCAGGGTCCAG 28  
DB 1762 CCATCGCAGGGTCCAG 1746

RESULT 6  
US-09-659-786-200/c  
;/ Sequence 200, Application US/09659786  
;/ Patent No. 6491894  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Ruoslahti, Erkki  
;/ APPLICANT: Pasqualini, Renata  
;/ TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
;/ TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
;/ TITLE OF INVENTION: Same  
;/ FILE REFERENCE: P-LU 3203  
;/ CURRENT APPLICATION NUMBER: US/09/659,786  
;/ CURRENT FILING DATE: 2000-09-11  
;/ PRIOR APPLICATION NUMBER: 08/926,914  
;/ PRIOR FILING DATE: 1997-09-10  
;/ PRIOR APPLICATION NUMBER: 08/710,067  
;/ PRIOR FILING DATE: 1996-09-10  
;/ NUMBER OF SEQ ID NOS: 226

;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 200  
;/ LENGTH: 3494  
;/ TYPE: DNA  
;/ ORGANISM: Homo sapiens  
;/ FEATURE:  
;/ NAME/KEY: CDS  
;/ LOCATION: (121)..(3024)  
US-09-659-786-200

Query Match 13.7%; Score 17; DB 4; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CCATCGCAGGGTCCAG 28  
DB 1762 CCATCGCAGGGTCCAG 1746

RESULT 7  
US-08-961-527-98/c  
;/ Sequence 96, Application US/08961527  
;/ Patent No. 6420135  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Charles Kunsch  
;/ TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
;/ NUMBER OF SEQUENCES: 391  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Human Genome Sciences, Inc.  
;/ STREET: 9410 Key West Avenue  
;/ CITY: Rockville  
;/ STATE: Maryland  
;/ COUNTRY: USA  
;/ ZIP: 20850  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
;/ COMPUTER: HP Vectra 486/33  
;/ OPERATING SYSTEM: MSDOS version 6.2  
;/ SOFTWARE: ASCII Text  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/961,527  
;/ CLASSIFICATION: 424  
;/ FILING DATE:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER:  
;/ FILING DATE:  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Brookes, A. Anders  
;/ REGISTRATION NUMBER: 36,373  
;/ REFERENCE/DOCKET NUMBER: PB340P1  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (301) 309-8504  
;/ TELEFAX: (301) 309-8512  
;/ INFORMATION FOR SEQ ID NO: 98:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 8654 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
US-08-961-527-98

Query Match 13.7%; Score 17; DB 4; Length 8654;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 GACTGTGTGTGTCTG 90  
DB 1105 GACTGTGTGTGTCTG 1089

RESULT 8  
US-09-313-294A-3641/c  
;/ Sequence 3641, Application US/09313294A

```
/ Patent No. 6476212
/ GENERAL INFORMATION:
/ APPLICANT: Lalgudi, Raghunath V.
/ APPLICANT: Ito, Laura Y.
/ APPLICANT: Sherman, Bradley K.
/ TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
/ FILE REFERENCE: PL-0017 US
/ CURRENT APPLICATION NUMBER: US/09/313,294A
/ CURRENT FILING DATE: 1999-05-14
/ NUMBER OF SEQ ID NOS: 7600
/ SOFTWARE: PERL Program
/ SEQ ID NO 3641
/ LENGTH: 266
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. 6476212 700612036H1
/ LOCATION: 128
/ OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-3641

Query Match          12.9%; Score 16; DB 4; Length 266;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      47 AGATGTCAGCTGAAA 62
      |||||
Db      155 AGATGTCAGCTGAAA 140

RESULT 9
US-09-149-476-239/c
/ Sequence 239, Application US/09149476
/ Patent No. 6420526
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 186 Human Secreted proteins
/ FILE REFERENCE: P2002P1
/ CURRENT APPLICATION NUMBER: US/09/149,476
/ CURRENT FILING DATE: 1998-09-08
/ EARLIER APPLICATION NUMBER: PCT/US98/04493
/ EARLIER FILING DATE: 1998-03-06
/ EARLIER APPLICATION NUMBER: 60/040,162
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,333
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/038,621
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,626
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,334
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,336
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,163
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/047,600
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,615
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,597
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,502
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,633
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,583
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,617
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,618
```

```
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,503
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,592
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,581
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,584
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,500
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,587
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,492
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,598
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,613
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,582
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,596
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,612
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,632
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,601
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,580
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,568
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,314
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,569
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,311
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,671
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,674
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,669
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,312
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,313
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,672
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,315
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/048,974
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/056,886
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,877
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,889
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,893
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,630
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,878
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,662
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,872
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,882
/ EARLIER FILING DATE: 1997-08-22
```

EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,539  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,553  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 12.9%; Score 16; DB 4; Length 458;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TCTGCAGGTCACGT 30  
DB 238 TCTGCAGGTCACGT 223

RESULT 10  
US-09-149-476-76/c  
Sequence 76, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23

[illegible]

EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 12.9%; Score 16; DB 4; Length 475;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TCTGCGAGGTCGAGT 30  
DB 226 TCTGCGAGGTCGAGT 211

RESULT 11

US-08-163-919A-1  
Sequence 1, Application US/08163919A

GENERAL INFORMATION:  
APPLICANT: THOMAS, Wayne R., CHUA, Kaw-Yan, ROGERS, Bruce L., and  
APPLICANT: KUO, Mei-chang  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING A HOUSE DUST MITE  
TITLE OF INVENTION: ALLERGEN, DER P III, AND USES THEREFOR  
NUMBER OF SEQUENCES: 18  
CURRENT APPLICATION DATA:  
FILING DATE: 08-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IPC-072 (IMI-041)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1059 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 63..848  
NAME/KEY: mat\_peptide  
LOCATION: 150..848  
US-08-163-919A-1

Query Match 12.9%; Score 16; DB 3; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 AGCTGAAACAAACAA 70  
DB 958 AGCTGAAACAAACAA 973

RESULT 12

PCT-US94-14073-1  
Sequence 1, Application PC/TUS9414073

GENERAL INFORMATION:  
APPLICANT: THOMAS, Wayne R., CHUA, Kaw-Yan, ROGERS, Bruce L., and  
APPLICANT: KUO, Mei-chang  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING A HOUSE DUST MITE  
TITLE OF INVENTION: ALLERGEN, DER P III, AND USES THEREFOR  
NUMBER OF SEQUENCES: 18  
CURRENT APPLICATION DATA:  
FILING DATE: 08-DEC-1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IPC-072 (IMI-041)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1059 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 63..848  
NAME/KEY: mat\_peptide  
LOCATION: 150..848  
PCT-US94-14073-1

Query Match 12.9%; Score 16; DB 5; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 AGCTGAAACAAACAA 70  
DB 958 AGCTGAAACAAACAA 973

RESULT 13

US-08-652-971-1/C  
Sequence 1, Application US/08652971

GENERAL INFORMATION:  
APPLICANT: Cheng, Jili  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE

TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,971  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 952-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-652-971-1

Query Match 12.9%; Score 16; DB 1; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAGGAGA 96  
|||||  
DB 890 TTGTGCTGGGAGGAGA 875

RESULT 14  
US-08-991-258A-1/c  
Sequence 1, Application US/08991258A  
Patent No. 592887  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,258A  
FILING DATE: 17-DEC-1997  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63478-3/WHD/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-991-258A-1

Query Match 12.9%; Score 16; DB 2; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAGGAGA 96  
|||||  
DB 890 TTGTGCTGGGAGGAGA 875

RESULT 15  
US-08-769-399-1/c  
Sequence 1, Application US/08769399  
Patent No. 5976852  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,399  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 225-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)



FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-769-399-1

Query Match 12.9%; Score 16; DB 2; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAG 96  
|||||  
Db 890 TTGTGCTGGGAG 875

RESULT 16  
US-08-991-953A-1/C

Sequence 1, Application US/08991953A  
Patent No. 6083748

GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, FTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,953A  
FILING DATE: 16-DEC-1997  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63478-3/WH/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX:

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-991-953A-1

Query Match 12.9%; Score 16; DB 3; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAG 96  
|||||  
Db 890 TTGTGCTGGGAG 875

RESULT 17

US-08-663-082-1  
Sequence 1, Application US/08663082  
Patent No. 6043411

GENERAL INFORMATION:  
APPLICANT: NISHIZAWA, Osamu  
TITLE OF INVENTION: GENE FOR FATTY ACID DESATURASE, VECTOR  
TITLE OF INVENTION: CONTAINING SAID GENE, PLANT TRANSFORMED WITH SAID GENE,  
TITLE OF INVENTION: AND PROCESS FOR CREATING SAID PLANT  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663,082  
FILING DATE: 25-JUN-1996  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP94/02288  
FILING DATE: 28-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 93/352858  
FILING DATE: 28-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 81356/107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
ORGANISM: Anabaena variabilis  
STRAIN: IAM M-3  
US-08-663-082-1

Query Match 12.1%; Score 15; DB 3; Length 196;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAG 95  
|||||  
Db 41 TTGTGCTGGGAG 55

RESULT 18  
US-09-313-294A-2588/C

Sequence 2588, Application US/09313294A  
Patent No. 6476212

GENERAL INFORMATION:  
APPLICANT: Lalugudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600

SOFTWARE: PERL Program  
SEQ ID NO: 2588  
LENGTH: 275  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Inocyte ID No. 6476212 700552942H1  
US-09-313-294A-2588

Query Match 12.1%; Score 15; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 CAAGAGACTGTGTT 82  
DB 178 CAAGAGACTGTGTT 164

RESULT 19  
US-08-309-182B-2  
Sequence 2, Application US/08309182B  
Patent No. 5639645  
GENERAL INFORMATION:  
APPLICANT: No. 563964510 MURATA  
TITLE OF INVENTION: A RECOMBINANT 9 DESATURASE AND A GENE  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/309,182B  
FILING DATE: September 20, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 819 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Anabaena variabilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..816  
IDENTIFICATION METHOD: P  
US-08-309-182B-2

Query Match 12.1%; Score 15; DB 1; Length 819;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 TTGTGCTGGGAG 95  
DB 521 TTGTGCTGGGAG 535

RESULT 20  
US-09-252-991A-11847  
Sequence 11847, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11847  
LENGTH: 942  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11847

Query Match 12.1%; Score 15; DB 4; Length 942;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GGTGATGATGCTTG 42  
DB 489 GGTGATGATGCTTG 503

RESULT 21  
US-09-252-991A-11595/c  
Sequence 11595, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11595  
LENGTH: 978  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11595

Query Match 12.1%; Score 15; DB 4; Length 978;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GGTGATGATGCTTG 42  
DB 424 GGTGATGATGCTTG 410

RESULT 22  
US-09-252-991A-5529/c  
Sequence 5529, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5529  
LENGTH: 1137  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1131)  
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-5529

Query Match 12.1%; Score 15; DB 4; Length 1137;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TGGCCATCTGCAGG 22  
DB 813 TGGCCATCTGCAGG 799

## RESULT 23

US-08-290-448A-71  
Sequence 71, Application US/08290448A  
Patent No. 5676954

GENERAL INFORMATION:  
APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Ratnar, Thorunn  
APPLICANT: Kuo, Mei-chang  
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,448A  
FILING DATE: August 15, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1196 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1161  
US-08-290-448A-71

Query Match 12.1%; Score 15; DB 1; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTATCATTC 108  
DB 449 AGAATGTATCATTC 463

RESULT 24  
US-08-290-448A-71  
Sequence 71, Application US/08290448A  
Patent No. 5698204

GENERAL INFORMATION:  
APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Ratnar, Thorunn  
APPLICANT: Kuo, Mei-chang  
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,448A  
FILING DATE: August 15, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1196 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1161  
US-08-290-448A-71

Query Match 12.1%; Score 15; DB 1; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTATCATTC 108  
DB 449 AGAATGTATCATTC 463

```
RESULT 25
US-08-175-069A-71
; Sequence 71, Application US/08175069A
; Patent No. 5776761
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,069A
; FILING DATE: December 29, 1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/529,951
; FILING DATE: May 29, 1990
; APPLICATION NUMBER: US 07/325,365
; FILING DATE: March 17, 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IM1-018DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1161
US-08-175-069A-71

Query Match      12.1%; Score 15; DB 1; Length 1196;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      94 AGAATGTAATCATTC 108
DB      449 AGAATGTAATCATTC 463

RESULT 26
US-08-461-939B-71
; Sequence 71, Application US/08461939B
; Patent No. 6335019
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Methods For Treating Sensitivity To A
; TITLE OF INVENTION: Protein Allergen Using Peptides Which Include A T Cell Epitope
; NUMBER OF SEQUENCES: 93
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,939B
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/464,000
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: US 08/290,448
; FILING DATE: 15-AUG-1994
; APPLICATION NUMBER: US 07/529,951
; FILING DATE: 29-MAY-1990
; APPLICATION NUMBER: US 07/325,365
; FILING DATE: 17-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IM1-018CNDV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1161
US-08-461-939B-71

Query Match      12.1%; Score 15; DB 4; Length 1196;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      94 AGAATGTAATCATTC 108
DB      449 AGAATGTAATCATTC 463

RESULT 27
US-08-464-000-71
; Sequence 71, Application US/08464000
; Patent No. 6335020
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,000  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/290,448  
FILING DATE: 15-AUG-1994  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: 29-MAY-1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: 17-MAR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMT-018CN2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1196 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1161  
US-08-464-000-71

Query Match 12.1% Score 15; DB 4; Length 1196;  
Best Local Similarity 100.0%; Pred.No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 AGAATGTAATCATTC 108  
|||||||  
Db 449 AGAATGTAATCATTC 463

RESULT 28  
US-09-205-258-189  
Sequence 189, Application US/09205258  
Patent No. 6525174  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007P1  
CURRENT APPLICATION NUMBER: US/09/205,258  
FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971

EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 189  
LENGTH: 1292  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-205-258-189

Query Match 12.1% Score 15; DB 4; Length 1292;  
Best Local Similarity 100.0%; Pred.No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 CTGAACAACAACAG 71  
|||||||  
Db 60 CTGAACAACAACAG 74

RESULT 29  
US-09-025-059-2  
; Sequence 2, Application US/09025059  
; Patent No. 6075136  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Giegler, Karl J.  
; TITLE OF INVENTION: PROSTATE-ASSOCIATED SERINE PROTEASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/025,059  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0481 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1314 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LUNGTUT10  
; CLONE: 2723646  
; US-09-025-059-2

Query Match 12.1%; Score 15; DB 3; Length 1314;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 CTGAACAACAACAG 71  
|||||  
DB 77 CTGAACAACAACAG 91

RESULT 30  
US-08-290-448A-79  
; Sequence 79, Application US/08290448A  
; Patent No. 5676954  
; GENERAL INFORMATION:  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Klapper, David G.  
; APPLICANT: Ratnar, Thorunn  
; APPLICANT: Kuo, Mei-chang  
; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: Boston

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,448A  
FILING DATE: August 15, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Maddagouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1368 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1366  
US-08-290-448A-79

Query Match 12.1%; Score 15; DB 1; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTATCATTC 108  
|||||  
DB 452 AGAATGTATCATTC 466

RESULT 31  
US-08-290-448A-79  
; Sequence 79, Application US/08290448A  
; Patent No. 5698204  
; GENERAL INFORMATION:  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Klapper, David G.  
; APPLICANT: Ratnar, Thorunn  
; APPLICANT: Kuo, Mei-chang  
; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,448A  
; FILING DATE: August 15, 1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/529,951  
; FILING DATE: May 29, 1990

APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ. ID NO.: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1368 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1366  
US-08-290-448A-79

Query Match 12.1%; Score 15; DB 1; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTAATCATTC 108  
DB 452 AGAATGTAATCATTC 466

RESULT 32  
US-08-175-069A-79  
Sequence 79, Application US/08175069A  
Patent No. 5776761  
GENERAL INFORMATION:  
APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Rafnar, Thorunn  
APPLICANT: Kuo, Mei-chang  
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/175,069A  
FILING DATE: December 29, 1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ. ID NO.: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1368 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1366  
US-08-175-069A-79

Query Match 12.1%; Score 15; DB 1; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTAATCATTC 108  
DB 452 AGAATGTAATCATTC 466

RESULT 33  
US-08-461-939B-79  
Sequence 79, Application US/08461939B  
Patent No. 6335019  
GENERAL INFORMATION:  
APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Rafnar, Thorunn  
APPLICANT: Kuo, Mei-chang  
TITLE OF INVENTION: Methods For Treating Sensitivity To A  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,939B  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/464,000  
FILING DATE: 05-JUN-1995  
APPLICATION NUMBER: US 08/290,448  
FILING DATE: 15-AUG-1994  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: 29-MAY-1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: 17-MAR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CNDV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)742-4214  
TELEFAX: (617)227-7400  
INFORMATION FOR SEQ. ID NO.: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1368 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1366  
US-08-461-939B-79

Query Match 12.1%; Score 15; DB 4; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 71;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTAATCATTC 108  
 |||||  
 DB 452 AGAATGTAATCATTC 466

RESULT 34  
 US-08-464-000-79

; Sequence 79, Application US/08464000

; Patent No. 6335020

; GENERAL INFORMATION:

; APPLICANT: Rogers, Bruce

; APPLICANT: Klapner, David G.

; APPLICANT: Rafnar, Thorunn

; APPLICANT: Kuo, Mei-chang

; TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD, LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/464,000

; FILING DATE: 05-JUN-1995

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/290,448

; FILING DATE: 15-AUG-1994

; APPLICATION NUMBER: US 07/529,951

; FILING DATE: 29-MAY-1990

; APPLICATION NUMBER: US 07/325,365

; FILING DATE: 17-MAR-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Amy E. Mandragoras

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: INT-018CN2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1368 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1366

; US-08-464-000-79

Query Match 12.1%; Score 15; DB 4; Length 1368;

Best Local Similarity 100.0%; Pred. No. 71;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGAATGTAATCATTC 108  
 |||||  
 DB 452 AGAATGTAATCATTC 466

RESULT 35  
 US-09-121-425-3

; Sequence 3, Application US/09121425

; Patent No. 6153418

; GENERAL INFORMATION:

; APPLICANT: Lehmann, Martin

; TITLE OF INVENTION: Consensus Phytases

; FILE REFERENCE: consensus phytases 13239

; CURRENT APPLICATION NUMBER: US/09/121,425

; CURRENT FILING DATE: 1998-07-23

; EARLIER APPLICATION NUMBER: EPO 97112688.3

; EARLIER FILING DATE: 1997-07-24

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO: 3

; LENGTH: 1426

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:consensus

; US-09-121-425-3

Query Match 12.1%; Score 15; DB 3; Length 1426;

Best Local Similarity 100.0%; Pred. No. 70;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ATGTCAAGCTGAAAA 63  
 |||||  
 DB 1250 ATGTCAAGCTGAAAA 1264

RESULT 36  
 US-09-634-493A-3

; Sequence 3, Application US/09634493A

; Patent No. 6579975

; GENERAL INFORMATION:

; APPLICANT: Lehmann, Martin

; TITLE OF INVENTION: Consensus Phytases

; FILE REFERENCE: consensus phytases 13239

; CURRENT APPLICATION NUMBER: US/09/634,493A

; CURRENT FILING DATE: 2000-08-08

; PRIOR APPLICATION NUMBER: US/09/121,425

; PRIOR FILING DATE: 1998-07-23

; PRIOR APPLICATION NUMBER: EPO 97112688.3

; PRIOR FILING DATE: 1997-07-24

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO: 3

; LENGTH: 1426

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:consensus

; US-09-634-493A-3

Query Match 12.1%; Score 15; DB 4; Length 1426;

Best Local Similarity 100.0%; Pred. No. 70;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ATGTCAAGCTGAAAA 63  
 |||||  
 DB 1250 ATGTCAAGCTGAAAA 1264

RESULT 37  
 US-08-149-105-4/C

; Sequence 4, Application US/08149105

; Patent No. 5538892

; GENERAL INFORMATION:

; APPLICANT: Donahoe, Patricia K.

; APPLICANT: Gustafson, Michael

; APPLICANT: He, Wei W.

; APPLICANT: Wang, Xiao-Fan

; TITLE OF INVENTION: TGF- TYPE I RECEPTOR

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:



```

; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,105
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/029,673
; FILING DATE: March 11, 1993
; APPLICATION NUMBER: 07/853,396
; FILING DATE: March 18, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/211001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1506
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-149-105-4
;
Query Match 12.1%; Score 15; DB 1; Length 1506;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ATCATTCTTCCACA 116
DB 902 ATCATTCTTCCACA 888

RESULT 38
US-08-317-847-4/c
; Sequence 4, Application US/08317847
; Patent No. 5547854
; GENERAL INFORMATION:
; APPLICANT: Donahoe, Patricia K.
; APPLICANT: Gustafson, Michael
; TITLE OF INVENTION: FOUR NOVEL RECEPTORS OF THE TGF-B
; TITLE OF INVENTION: FAMILY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,847
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 08/029,673
; FILING DATE: March 11, 1993
; APPLICATION NUMBER: 07/853,396
; FILING DATE: March 18, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/127002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1506
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-317-847-4
;
Query Match 12.1%; Score 15; DB 1; Length 1506;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ATCATTCTTCCACA 116
DB 902 ATCATTCTTCCACA 888

RESULT 39
US-09-252-991A-5553/c
; Sequence 5553, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5553
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-5553
;
Query Match 12.1%; Score 15; DB 4; Length 1545;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TGGCCATCTGCAGG 22
DB 1280 TGGCCATCTGCAGG 1266

RESULT 40
US-09-252-991A-11780
; Sequence 11780, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11780  
LENGTH: 1605  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11780

Query Match 12.1%; Score 15; DB 4; Length 1605;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 GGTGATGATGCTG 42  
|||||  
Db 409 GGTGATGATGCTG 423

## RESULT 41

US-08-123-934A-9/C  
Sequence 9, Application US/08123934A  
Patent No. 6291206  
GENERAL INFORMATION:  
APPLICANT: MOZNEY, John  
APPLICANT: CELESTE, Anthony J.  
APPLICANT: THIES, R. Scott  
APPLICANT: YAMAUTI, No. 62912060ru  
TITLE OF INVENTION: RECEPTOR PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute Inc.- Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,934A  
FILING DATE: 17-SEP-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: LAZAR, Steven R  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5203  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 876 1170  
TELEFAX: 617 876 5851  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1794 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: W-120  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 83..1591  
US-08-123-934A-9

Query Match 12.1%; Score 15; DB 3; Length 1794;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 ATCATCTCTCCACA 116  
|||||  
Db 990 ATCATCTCTCCACA 976

## RESULT 42

PCT-US94-10080-9/C  
Sequence 9, Application PC/TUS9410080  
GENERAL INFORMATION:  
APPLICANT: GENETICS INSTITUTE, INC.  
TITLE OF INVENTION: RECEPTOR PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute Inc.- Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/10080  
FILING DATE: HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,934  
FILING DATE: 17-SEP-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: LAZAR, Steven R  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5203-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-5851  
TELEFAX: (617) 498-8260  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1794 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: W-120  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 83..1591  
PCT-US94-10080-9

Query Match 12.1%; Score 15; DB 5; Length 1794;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 ATCATCTCTCCACA 116  
|||||  
Db 990 ATCATCTCTCCACA 976

## RESULT 43

US-09-252-991A-5670  
Sequence 5670, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
RILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5670  
LENGTH: 1914  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5670

Query Match 12.1%; Score 15; DB 4; Length 1914;  
Best Local Similarity 100.0%; Pred.No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TGGCCATCTGCAGG 22  
DB 209 TGGCCATCTGCAGG 223

RESULT 44  
US-09-382-256-9/c  
Sequence 9, Application US/09382256A  
Patent No. 6207814  
GENERAL INFORMATION:  
APPLICANT: MIYAZONO, Kohei  
FRANZEN, Petra  
YAMASHITA, Hideoshi  
HELDIN, Carl-Henrik

TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS  
HAVING SERINE THREONINE KINASE DOMAINS,  
AND THEIR USE

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSER: Fulbright & Jaworski L.L.P.

STREET: 666 Fifth Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/382,256A

FILING DATE: 24-Aug-1999

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/02367

FILING DATE: No. 6207814ember 17, 1993

APPLICATION NUMBER: GB 9224057.1

FILING DATE: No. 6207814ember 17, 1992

APPLICATION NUMBER: GB 9304677.9

FILING DATE: March 8, 1993

APPLICATION NUMBER: GB 9304680.3

FILING DATE: March 8, 1993

APPLICATION NUMBER: 9311047.6

FILING DATE: May 28, 1993

APPLICATION NUMBER: 9313763.6

FILING DATE: July 2, 1993

APPLICATION NUMBER: 9316099.2

FILING DATE: August 3, 1993

APPLICATION NUMBER: 321344.5

FILING DATE: October 15, 1993

ATTORNEY/AGENT INFORMATION:

NAME: No. 6207814man D. Hanson

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5298.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3000

TELEFAX: (212) 752-5958

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 2308 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:

ORGANISM: Mouse  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 77..1585

US-09-382-256-9  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Query Match 12.1%; Score 15; DB 3; Length 2308;  
Best Local Similarity 100.0%; Pred.No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ATCATTCTTCACA 116  
DB 984 ATCATTCTTCACA 970

RESULT 45  
US-09-395-115-9/c  
Sequence 9, Application US/09395115  
Patent No. 6271365  
GENERAL INFORMATION:

APPLICANT: Miyazono, Kohei; Franzen, Peter Ten;

YAMASHITA, Hideoshi; Heldin, Carl-Henrik

TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins And Their Use

TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSER: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 Kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/395,115

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/436,265

FILING DATE: 30-October-1995

APPLICATION NUMBER: PCT/GB93/02367

FILING DATE: 17-No. 6271365ember-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9224057.1

FILING DATE: 17-No. 6271365ember-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304677.9

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304680.3

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9311047.6

FILING DATE: 28-May-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9313763.6

FILING DATE: 2-July-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9136099.2

FILING DATE: 3-August-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321344.5  
FILING DATE: 15-October-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohlei, Vineet  
REGISTRATION NUMBER: 37,003  
REFERENCE/DOCKET NUMBER: LUD 5298  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2308 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE: Mouse  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 77..1585  
US-09-395-115-9

Query Match 12.1%; Score 15; DB 3; Length 2308;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 ATCATTCTTCACCA 116  
Db 984 ATCATTCTTCACCA 970

RESULT 46  
US-08-436-265-9/c  
Sequence 9, Application US/08436265  
Patent No. 6316217  
GENERAL INFORMATION:  
APPLICANT: Miyazono, Kohlei, DiJke, Peter Ten;  
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik  
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins  
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,265  
FILING DATE: 30-October-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/02367  
FILING DATE: 17-No. 6316217ember-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9224057.1  
FILING DATE: 17-No. 6316217ember-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9304677.9  
FILING DATE: 8-March-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9304680.3  
FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9311047.6  
FILING DATE: 28-May-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9313763.6  
FILING DATE: 2-July-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9136099.2  
FILING DATE: 3-August-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321344.5  
FILING DATE: 15-October-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohlei, Vineet  
REGISTRATION NUMBER: 37,003  
REFERENCE/DOCKET NUMBER: LUD 5298  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2308 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE: Mouse  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 77..1585  
US-08-436-265-9

Query Match 12.1%; Score 15; DB 4; Length 2308;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 ATCATTCTTCACCA 116  
Db 984 ATCATTCTTCACCA 970

RESULT 47  
US-09-679-187-9/c  
Sequence 9, Application US/09679187  
Patent No. 6331621  
GENERAL INFORMATION:  
APPLICANT: Miyazono, Kohlei, DiJke, Peter Ten;  
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik  
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins  
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/679,187  
FILING DATE: 03-OCT-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,265  
FILING DATE: 30-October-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/02367  
FILING DATE: 17-No. 6331621ember-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9224057.1  
FILING DATE: 17-No. 6331621ember-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9304677.9  
FILING DATE: 8-March-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9304680.3  
FILING DATE: 8-March-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9311047.6  
FILING DATE: 28-May-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9313763.6  
FILING DATE: 2-July-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9136099.2  
FILING DATE: 3-August-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321344.5  
FILING DATE: 15-October-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohler, Vineet  
REGISTRATION NUMBER: 37,003  
REFERENCE/DOCKET NUMBER: LUD 5298  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2308 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Mouse  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 77..1585  
US-09-679-187-9

Query Match 12.1%; Score 15; DB 4; Length 2308;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ATCATTCTTCCACA 116  
DB 984 ATCATTCTTCCACA 970

RESULT 48  
US-09-489-847-67  
Sequence 67, Application US/09489847  
GENERAL INFORMATION:  
APPLICANT: Rosen et al  
TITLE OF INVENTION: 98 Human Secreted Proteins  
FILE REFERENCE: P2031P1  
CURRENT APPLICATION NUMBER: US/09/489,847  
EARLIER APPLICATION NUMBER: PCT/US99/17130  
EARLIER FILING DATE: 1999-07-29  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
EARLIER APPLICATION NUMBER: 60/095,486  
EARLIER FILING DATE: 1998-08-05  
EARLIER APPLICATION NUMBER: 60/096,319

EARLIER FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: 60/095,454  
EARLIER FILING DATE: 1998-08-06  
EARLIER APPLICATION NUMBER: 60/095,455  
EARLIER FILING DATE: 1998-08-06  
NUMBER OF SEQ ID NOS: 376  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 67  
LENGTH: 2434  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (10)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (12)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (27)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (73)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (75)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (103)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (130)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-489-847-67

Query Match 12.1%; Score 15; DB 4; Length 2434;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 CTTAGATGTCAAGCT 58  
DB 31 CTTAGATGTCAAGCT 45

RESULT 49  
US-07-959-943-10/c  
Sequence 10, Application US/07959943  
Patent No. 5418162  
GENERAL INFORMATION:  
APPLICANT: Blakely, Randy D.  
APPLICANT: Fremieu Jr., Robert T.  
TITLE OF INVENTION: Serotonin Transporter cDNA  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley, Bell, Seltzer, Park  
ADDRESSER: and  
ADDRESSEE: Gibson  
STREET: Post Office Drawer 31107  
CITY: Raleigh  
STATE: No. 5418162th Carolina  
COUNTRY: U.S.A.  
ZIP: 27622  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/959,943
CLASSIFICATION: 435
FILING DATE: 19921014
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405.38a
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-881-3140
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2508 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
POSITION IN GENOME:
UNITS: 2278 basepairs
FEATURE:
NAME/KEY: CDS
LOCATION: 73..1962
US-07-959-943-10

```

```

Query Match      12.1%; Score 15; DB 1; Length 2508;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      20 AGGGTCACGTGATG 34
Db      737 AGGGTCACGTGATG 723

```

```

RESULT 50
US-08-551-459-3/c
Sequence 3, Application US/08551459
Patent No. 5821350
GENERAL INFORMATION:
APPLICANT: Huang, Yue
APPLICANT: Karatzas, Costas N.
APPLICANT: Lazaris-Karatzas, Anthoula
APPLICANT: Delaquis, Annick
TITLE OF INVENTION: ASPERGILLUS NIGER BETA GALACTOSIDASE
TITLE OF INVENTION: GENE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/551,459
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06632/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-8906
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3057 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
US-08-551-459-3

```

```

Query Match      12.1%; Score 15; DB 1; Length 3057;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      47 AGATGTCACCTGAA 61
Db      248 AGATGTCACCTGAA 234

```

Search completed: November 7, 2003, 11:59:34  
Job time : 34.5936 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 11:11:37 / Search time 842.956 Seconds

(without alignments)  
6163.455 Million cell updates/sec

Title: US-09-509-779-1\_COPY\_154\_280

Perfect score: 127

Sequence: 1 GTGGCATACCTGTGCCATCT.....ATTCTTCACACACTGTCTGC 127

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: GenEmbl1:

```

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	127	100.0	1125	10	BC011127	BC011127 Mus muscu
2	127	100.0	1140	6	BD096966	BD096966 SAG:apopt
3	127	100.0	1140	10	AF092877	AF092877 Mus muscu
4	84	66.1	219370	10	AC123935	AC123935 Mus muscu
5	69	54.3	120515	10	AL929565	AL929565 Mouse DNA
6	49	38.6	244851	2	AC118615	AC118615 Mus muscu
7	38	29.9	612	11	BW071691	BW071691 S212P6546
8	34	26.8	112321	2	AC103666	AC103666 Mus muscu
9	33	26.0	187353	2	AC118192	AC118192 Mus muscu
10	32	25.2	224	6	BD077705	BD077705 5'EST of
11	32	25.2	342	6	AX282592	AX282592 Sequence
12	32	25.2	342	9	AF142060	AF142060 Homo sapi
13	32	25.2	342	9	BT007348	BT007348 Homo sapi
14	32	25.2	754	6	BD096967	BD096967 SAG:apopt
15	32	25.2	754	6	BD096981	BD096981 SAG:apopt
16	32	25.2	754	6	BD096985	BD096985 SAG:apopt
17	32	25.2	754	6	BD096986	BD096986 SAG:apopt
18	32	25.2	754	6	BD096987	BD096987 SAG:apopt
19	32	25.2	754	6	BD096988	BD096988 SAG:apopt
20	32	25.2	754	6	BD096989	BD096989 SAG:apopt
21	32	25.2	754	6	BD096990	BD096990 SAG:apopt
22	32	25.2	754	6	BD096991	BD096991 SAG:apopt
23	32	25.2	754	6	BD096992	BD096992 SAG:apopt
24	32	25.2	754	6	BD096994	BD096994 SAG:apopt
25	32	25.2	754	6	BD096995	BD096995 SAG:apopt
26	32	25.2	754	9	AF092878	AF092878 Homo sapi
27	32	25.2	816	9	BC008627	BC008627 Homo sapi
28	32	25.2	822	9	BC005966	BC005966 Homo sapi
29	32	25.2	836	9	AF164679	AF164679 Homo sapi
30	32	25.2	187353	2	AC118192	AC118192 Mus muscu
31	31	24.4	254832	2	AC095698	AC095698 Rattus no
32	31	23.6	754	6	BD096984	BD096984 SAG:apopt
33	31	23.6	754	6	BD096993	BD096993 SAG:apopt
34	30	23.6	148290	2	AC119145	AC119145 Rattus no
35	30	23.6	250425	2	AC125667	AC125667 Rattus no
36	29	22.8	674	9	AF312226	AF312226 Homo sapi
37	29	22.8	747	6	BD096973	BD096973 SAG:apopt
38	29	22.8	754	6	BD096982	BD096982 SAG:apopt
39	29	22.8	754	6	BD096983	BD096983 SAG:apopt
40	27	21.3	112321	2	AC103666	AC103666 Mus muscu
41	25	19.7	439	9	HS323208	HS323208 Homo sapi
42	25	19.7	706	6	BD096974	BD096974 SAG:apopt
43	25	19.7	35638	9	AC112771	AC112771 Homo sapi
44	25	19.7	238330	2	AC106176	AC106176 Rattus no
45	25	19.7	245476	2	AC098496	AC098496 Rattus no

# ALIGNMENTS

RESULT 1  
BC011127  
LOCUS  
DEFINITION  
IMAGE:4194107), complete cds.  
ACCESSION  
BC011127  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
1 (bases 1 to 1125)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL
Altschul S.F., Zeeberg B., Buewoy K.H., Schaefer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T.T., Max S.I., Wang J., Hsieh F., Datchenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L., Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshitsuki S., Carinici P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W., Villalob D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A., Pathey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmittz J., Myers R.M., Butterfield V.S., Krzywinski M.I., Skalska U., Smilus D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	22388257	12477932	2 (bases 1 to 1125)	Straussberg R. Direct Submission Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgabs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LIML) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hnsc.bcm.tmc.edu/cdna/ Contact: amgcdm.tmc.edu Gunaratne P.H., Garcia, A.M., Lu, X., Huliy, S.W., Louised, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov Series: IRAX Plate: 24 Row X Column: 5 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, similarity but not identity to protein. Location/Qualifiers 1. .1125 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:19076 IMAGE:4194107" /tissue_type="Liver, normal, 5 month old male mouse." /clone_lib="NCI CGAP_L19" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" 1. .1125 /gene="Rnf7" /note="Synonym: SAG" /db_xref="locustID:19823" /db_xref="MGI:1337096" 12..353 /codon_start=1 /product="Rnf7 protein" /protein_id="AAH1127.1" /db_xref="GI:15029808" /translation="NADYDEGEPCVCYSSHGSGKSGSGGDYMPSLKKNNAVMSWDVECTICATCRQVMDACLRCAENKEQECVVWWECHNSFINCCMSLTWVKNNCEPLC

BASE COUNT	304 a	219 c	286 g	316 t
ORIGIN	OODWVVRIGK"			
Query Match	100.0%;	Score 127;	DB 10;	Length 1125;
Best Local Similarity	100.0%;	Pred. No. 5.1e-67;		
Matches 127; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0
Db	<div>1 GTGGCATTCCTGTGCATCTGACAGSGTCCAGSTGAATGCCTTGCTTCGATTGCCAATGCCAACG 60       149 GCGCATACCTGTGTCATCTGAGGGTCCAGGTGATGATGCTCCTTCGATGTCACG 208       61 TGAATAACAAGAAGAGACTGTGTGTGTGTGTGTGAGAGATGTAACCAATTCCTTCACAA 120 209 TGAATAACAAGAAGAGAGACTGTGTGTGTGTGTGTGAGAGATGTAACCAATTCCTTCACAA 268       121 CTGCTGC 127 269 CTGCTGC 275</div>			
Qy				
Db				
RESULT 2				
LOCUS	BD096966	1140 bp	DNA	linear
DEFINITION	SAG:apoptosis sensitivity gene.			
ACCESSION	BD096966	.1	GI:22642554	
VERSION	BD096966.1			
KEYWORDS	JP 2001526063-A/1.			
SOURCE	unidentified			
ORGANISM	unclassified			
REFERENCE	1 (bases 1 to 1140)			
AUTHORS	Sun, Y.			
TITLE	SAG:apoptosis sensitivity gene			
JOURNAL	Patent: JP 2001526063-A 1 18-DEC-2001; WARNER LAMBERT CO			
COMMENT	OS Unidentified PN JP 2001526063-A/1 PD 18-DEC-2001 PF 15-DEC-1998 JP 2000525451 PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI YI SUN PC C12N5/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00, PC A61P93/06, PC A61P93/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC C12N5/10,C12Q1/68, PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC Strandedness: Double; CC Topology: Linear; CC /note = 'Mouse SAG' FH Key Location/Qualifiers FT CDS 17..355 FT mat_peptide 17..355 FT misc_feature 1..1140. Location/Qualifiers 1..1140 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"			
FEATURES	source			
BASE COUNT	302 a	224 c	287 g	327 t
ORIGIN				
Query Match	100.0%;	Score 127;	DB 6;	Length 1140;
Best Local Similarity	100.0%;	Pred. No. 5.1e-67;		
Matches 127; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	<div>1 GTGGCATTCCTGTGCATCTGACAGSGTCCAGSTGAATGCCTTGCTTCGATTGCCAATGCCAACG 60       154 GTGGCATTCCTGTGCATCTGACAGSGTCCAGSTGAATGCCTTGCTTCGATTGCCAATGCCAACG 213       61 TGAATAACAAGAAGAGACTGTGTGTGTGTGTGTGAGAGATGTAACCAATTCCTTCACAA 120 214 TGAATAACAAGAAGAGACTGTGTGTGTGTGTGTGAGAGATGTAACCAATTCCTTCACAA 273      </div>			
Db				



QY 121 CTGCTGC 127  
 Db 274 CTGCTGC 280

RESULT 3  
 AF092877 1140 bp mRNA linear ROD 24-JUL-2001  
 LOCUS Mus musculus zinc RING finger protein SAG mRNA, complete cds.  
 DEFINITION AF092877  
 ACCESSION AF092877.1 GI:4588031  
 VERSION  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1140)  
 Duan,H., Wang,Y., Avitram,M., Swaroop,M., Loo,J.A., Bian,J.,  
 Tian,Y., Mueller,T., Bisgaier,C.L. and Sun,Y.  
 SAG, a novel zinc RING finger protein that protects cells from  
 apoptosis induced by redox agents  
 Mol. Cell. Biol. 19 (4), 3145-3155 (1999)

JOURNAL MEDLINE  
 PUBMED 99182502  
 REFERENCE 10082581  
 AUTHORS Sun, Y.  
 TITLE Alterations of SAG mRNA in human cancer cell lines: requirement for the RING finger domain for apoptosis protection  
 JOURNAL MEDLINE  
 PUBMED 99435944  
 REFERENCE 10506102  
 AUTHORS Swaroop,M., Bian,J., Avitram,M., Duan,H., Bisgaier,C.L., Loo,J.A. and Sun,Y.  
 TITLE Expression, purification, and biochemical characterization of SAG, a RING finger redox-sensitive protein  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS Swaroop,M., Wang,Y., Miller,P., Duan,H., Jackoe,T., Madore,S.J. and Sun,Y.  
 TITLE Yeast homolog of human SAG/ROCC/Rbx2/Hrt2 is essential for cell growth, but not for germination: chip profiling implicates its role in cell cycle regulation  
 JOURNAL MEDLINE  
 PUBMED 20309864  
 REFERENCE 10851089  
 AUTHORS Duan,H., Tsvetkov,L.M., Liu,Y., Song,Y., Swaroop,M., Wen,R., Kung,H.F., Zhang,H. and Sun,Y.  
 TITLE Promotion of S-phase entry and cell growth under serum starvation by SAG/ROCC/Rbx2/Hrt2, an E3 ubiquitin ligase component: association with inhibition of p27 accumulation  
 JOURNAL MEDLINE  
 PUBMED 21152847  
 REFERENCE 11255262  
 AUTHORS Mol. Carcinog. 30 (1), 37-46 (2001)  
 TITLE 6 (bases 1 to 1140)  
 Sun,Y.  
 JOURNAL MEDLINE  
 PUBMED 11255262  
 REFERENCE 11255262  
 AUTHORS Direct Submission  
 TITLE Submitted (16-SEP-1998) Department of Molecular Biology,  
 JOURNAL MEDLINE  
 PUBMED 11255262  
 REFERENCE 11255262  
 AUTHORS Parke-Davies, 2800 Plymouth Rd, Ann Arbor, MI 48105, USA  
 TITLE Location/Qualifiers  
 FEATURES  
 SOURCE 1..1140  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6xCBA"  
 /db\_xref="taxon:10090"  
 /sex="female"  
 /issue\_type="1ung"  
 /dev\_stage="6-8 weeks"  
 17..358  
 /note="redox sensitive, metal binding; expression protects cells from apoptosis induced by redox compounds"

BASE COUNT 302 a 224 c 287 g 327 t  
 ORIGIN

Query Match 100.0%; Score 127; DB 10; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 5,1e-67;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGATACCTGTCATGTCAGGTCAGGTGATGATGCTGCTTCGATGTCAC 60  
 Db 154 GGGGATACCTGTCATGTCAGGTCAGGTGATGATGCTGCTTCGATGTCAC 213

QY 61 TGAACAAGCAAG 120  
 Db 214 TGAACAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273

QY 121 CTGCTGC 127  
 Db 274 CTGCTGC 280

RESULT 4  
 AC123935 219370 bp DNA linear ROD 02-OCT-2002  
 LOCUS Mus musculus chromosome 13 clone RP23-354J3, complete sequence.  
 DEFINITION AC123935  
 ACCESSION AC123935  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 219370)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Unpublished  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Submitted (05-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Submitted (13-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Submitted (02-OCT-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL MEDLINE  
 PUBMED 10506102  
 REFERENCE 10506102  
 AUTHORS Parkway, St. Louis, MO 63108, USA  
 TITLE On Oct 2, 2002 this sequence version replaced gi:22748547.

COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.wustl.edu  
 Project Information  
 Center project name: M BA0354J03

FEATURES  
 source 1..219370  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="13"

BASE COUNT 61709 a 45568 c 45395 g 66698 t  
ORIGIN  
Query Match 66.1%; Score 84; DB 10; Length 219370;  
Best Local Similarity 100.0%; Pred. No. 2.1e-40;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGCATACCTGTGCTGAGGAGGTCAGGTGATGATGCTGCTGATGTCAGC 60  
DB 79010 GTGCATACCTGTGCTGAGGAGGTCAGGTGATGATGCTGCTGATGTCAGC 79069  
QY 61 TGAACAACAGCAGAGGACTGTGT 84  
DB 79070 TGAACAACAGCAGAGGACTGTGT 79093

RESULT 5  
AL929565/c 120515 bp DNA linear ROD 16-APR-2003  
LOCUS Mouse DNA sequence from clone RP23-454N16 on chromosome 4, complete  
ACCESSION AL929565  
VERSION AL929565  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 120515)  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (16-APR-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk  
On Apr 16, 2003 this sequence version replaced gi:29825605.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Emi: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-454N16 is  
from the RPI-23 Mouse BAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6.  
FEATURES  
SOURCE  
1..120515  
/organism="Mus musculus"  
/mol\_type="genomic DNA"

BASE COUNT 34369 a 25840 c 26441 g 33865 t  
ORIGIN  
Query Match 54.3%; Score 69; DB 10; Length 120515;  
Best Local Similarity 100.0%; Pred. No. 4.2e-31;  
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 47 CTTCGATGTCAGCTGAAACAGAGAGGAGCTGTGTGCTGGAGAGCTGTAAC 106  
DB 57827 CTTCGATGTCAGCTGAAACAGAGAGGAGCTGTGTGCTGGAGAGCTGTAAC 57768  
QY 107 CATTCCTTC 115  
DB 57767 CATTCCTTC 57759

RESULT 6  
AC118615/c 244851 bp DNA linear HTG 16-MAR-2003  
LOCUS Mus musculus clone RP24-91A16, WORKING DRAFT SEQUENCE, 7 unordered  
pieces.  
ACCESSION AC118615  
VERSION AC118615.4 GI:28975088  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 244851)  
REFERENCE  
AUTHORS Birren, B., Nussbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP24-91A16  
JOURNAL Unpublished  
2 (bases 1 to 244851)  
AUTHORS  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Camarata, A., Campolongo, A., Chang, J.,  
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Farro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Grinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,  
Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Punthang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Traviss, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, D., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (19-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02144, USA  
3 (bases 1 to 244851)  
REFERENCE  
AUTHORS  
Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Archcholi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavsky, L., Boukhalter, B., Camarata, A., Chang, J., Choepel, Y.,  
Collymore, A., Cook, A., Cooke, P., Corum, B., Deatellano, K.,  
Diaz, J.S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Farro, S.,  
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rie, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

Direct Submission  
Submitted (16-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2003 this sequence version replaced gi:28626759.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L24334  
Center clone name: 91.A.16

----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 243312 bases at least Q40  
Consensus quality: 243902 bases at least Q30  
Consensus quality: 244127 bases at least Q20  
Insert size: 244251; sum-of-contents  
Quality coverage: 9.5 in Q20 bases; sum-of-contents

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 76430: contig of 76430 bp in length  
\* 76431 76530: gap of 100 bp  
\* 76531 77709: contig of 1179 bp in length  
\* 77710 77809: gap of 100 bp  
\* 77810 78856: contig of 1047 bp in length  
\* 78857 78956: gap of 100 bp  
\* 78957 86294: contig of 7338 bp in length  
\* 86295 86394: gap of 100 bp  
\* 86395 96039: contig of 9645 bp in length  
\* 96040 96139: gap of 100 bp  
\* 96140 228506: contig of 132367 bp in length  
\* 228507 228606: gap of 100 bp  
\* 228607 224851: contig of 16245 bp in length.

Location/Qualifiers  
1. 244851  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone\_lib="RP24-91A16"  
/clone\_1lb="RP24-91A16"  
1. 76430  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
76531\_77709  
/note="assembly\_fragment"  
77810\_78856  
/note="assembly\_fragment"  
78957\_86294  
/note="assembly\_fragment"

misc\_feature 86395..96039 /note="assembly\_fragment"  
misc\_feature 96140..228506 /note="assembly\_fragment"  
misc\_feature 228607..224851 /note="assembly\_fragment"  
clone\_end:77  
vector\_side:right

BASE COUNT 66327 a 54365 c 54122 g 69435 t 602 others

ORIGIN  
Query Match 38.6%; Score 49; DB 2; Length 244851;  
Best Local Similarity 100.0%; Pred. No. 1e-18;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GATGCTGCTTCGATGTAAGCTGAAAAAGCAAGAGACTGTGTTG 86  
|||||  
DB 185253 GATGCTGCTTCGATGTAAGCTGAAAAAGCAAGAGACTGTGTTG 185205

RESULT 7  
BV071691  
LOCUS BV071691 612 bp DNA linear STS 31-MAY-2003  
DEFINITION S212P6546P3.70 CZECHII/Ei Mus musculus STS genomic, sequence  
tagged site.  
ACCESSION BV071691.1 GI:31187486  
VERSION BV071691  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (baaes 1 to 612)  
Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,  
Lander, E.S., Lindblad-Toh, K., and Daly, M.J.  
The mosaic structure of variation in the laboratory mouse genome  
Nature 420 (6915), 574-578 (2002)  
JOURNAL MEDLINE 22354684  
PUBMED 12466852

COMMENT  
Contact: Kerstin Lindblad-Toh  
Whitehead Institute for Biomedical Research, Center for Genome  
Research  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172521477  
Fax: 6172580903  
Email: kersli@genome.wi.mit.edu  
Primer A: None  
Primer B: None  
STS size: 612  
Protocol:  
WGS-discovery: Paired-end low-coverage whole genome shotgun reads  
were generated from 129S1/SVimJ, C3H/HeJ, and BALB/cByJ. The WGS  
reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP  
detection was carried out by SSHA-SNP. 225,000 reads were  
annotated  
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J  
and the strain from which the particular read came. The validation  
rate for these SNPs was estimated at approximately 98%.

Location/Qualifiers  
1. 612  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="CZECHII/Ei"  
/db\_xref="taxon:10090"  
/map="10 22-541 63463278-63462744"  
/clone\_1lb="CZECHII/Ei"

FEATURES  
source  
1. 612  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="CZECHII/Ei"  
/db\_xref="taxon:10090"  
/map="10 22-541 63463278-63462744"  
/clone\_1lb="CZECHII/Ei"

misc\_feature 145 a 163 c 170 g 134 t  
BASE COUNT  
ORIGIN  
Query Match 29.9%; Score 38; DB 1; Length 612;  
Best Local Similarity 100.0%; Pred. No. 7.6e-12;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

47 CTTGATGTCAGCTGAAACAGACAGAGACTGTGT 84  
 407 CTTGATGTCAGCTGAAACAGACAGAGACTGTGT 444

RESULT 8  
 AC103666 112321 bp DNA linear HTG 20-DEC-2002  
 LOCUS Mus musculus clone RP23-397H13, LOW-PASS SEQUENCE SAMPLING.  
 AC103666.2 GI:27311471  
 ACCESSION  
 VERSION  
 KEYWORDS HTG, HTGS\_PHASE0.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

DEFINITION Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 112321)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-397H13  
 Unpublished  
 2 (bases 1 to 112321)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barr, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Menus, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, W., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vasiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 112321)  
 AUTHORS Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barr, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vasiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Dec 20, 2002 this sequence version replaced gi:11719550.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)

http://fpc.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: U19615  
 Center Clone name: 397\_H\_13  
 -----  
 \* NOTE: This record contains 89 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1 1147: contig of 1147 bp in length  
 \* 1148 1247: gap of 100 bp  
 \* 1248 1248: contig of 1101 bp in length  
 \* 2349 2348: gap of 100 bp  
 \* 2449 2448: contig of 1063 bp in length  
 \* 3512 3511: gap of 100 bp  
 \* 3612 3611: gap of 100 bp  
 \* 4747 4746: contig of 1135 bp in length  
 \* 4846 4845: gap of 100 bp  
 \* 4847 4846: contig of 1184 bp in length  
 \* 6031 6030: contig of 100 bp  
 \* 6131 6130: gap of 100 bp  
 \* 6311 7290: contig of 1160 bp in length  
 \* 7291 7390: gap of 100 bp  
 \* 7391 7390: gap of 100 bp  
 \* 8578 8577: contig of 1187 bp in length  
 \* 8579 8578: gap of 100 bp  
 \* 8678 8677: gap of 100 bp  
 \* 9937 9936: contig of 1260 bp in length  
 \* 10037 10036: gap of 100 bp  
 \* 10038 11192: contig of 1155 bp in length  
 \* 11193 11292: gap of 100 bp  
 \* 11293 12454: contig of 1162 bp in length  
 \* 12455 12554: gap of 100 bp  
 \* 12555 13718: contig of 1164 bp in length  
 \* 13719 13818: gap of 100 bp  
 \* 13819 15002: contig of 1184 bp in length  
 \* 15003 15102: gap of 100 bp  
 \* 15103 16234: contig of 1132 bp in length  
 \* 16235 16334: gap of 100 bp  
 \* 16335 17461: contig of 1127 bp in length  
 \* 17462 17561: gap of 100 bp  
 \* 17562 18650: contig of 1089 bp in length  
 \* 18651 18750: gap of 100 bp  
 \* 18751 19912: contig of 1162 bp in length  
 \* 19913 20012: gap of 100 bp  
 \* 20013 21199: contig of 1187 bp in length  
 \* 21200 21299: gap of 100 bp  
 \* 21300 22462: contig of 1163 bp in length  
 \* 22463 22562: gap of 100 bp  
 \* 22563 23720: contig of 1158 bp in length  
 \* 23721 23820: gap of 100 bp  
 \* 23821 25005: contig of 1185 bp in length  
 \* 25006 25105: gap of 100 bp  
 \* 25106 26265: contig of 1160 bp in length  
 \* 26266 26365: gap of 100 bp  
 \* 26366 27528: contig of 1163 bp in length  
 \* 27529 27628: gap of 100 bp  
 \* 27629 28788: contig of 1160 bp in length  
 \* 28789 28888: gap of 100 bp  
 \* 28889 30034: contig of 1146 bp in length  
 \* 30035 30134: gap of 100 bp  
 \* 30135 31236: contig of 1102 bp in length  
 \* 31237 31336: gap of 100 bp  
 \* 31337 32512: contig of 1176 bp in length  
 \* 32513 32612: gap of 100 bp

32613	33913	config of 1201 bp in length
33814	33913	gap of 100 bp
33914	35107	config of 1194 bp in length
35108	35207	gap of 100 bp
35208	36431	config of 1224 bp in length
36432	36531	gap of 100 bp
36532	37701	config of 1170 bp in length
37702	37801	gap of 100 bp
37802	38967	config of 1166 bp in length
38968	39067	gap of 100 bp
39068	40235	config of 1166 bp in length
40236	40335	gap of 100 bp
40336	41492	config of 1157 bp in length
41493	41592	gap of 100 bp
41593	42691	config of 1099 bp in length
42692	42791	gap of 100 bp
42792	43970	config of 1179 bp in length
43971	44070	gap of 100 bp
44071	45150	config of 1080 bp in length
45151	45250	gap of 100 bp
45251	46347	config of 1097 bp in length
46348	46447	gap of 100 bp
46448	47608	config of 1161 bp in length
47609	47709	gap of 100 bp
47709	48881	config of 1177 bp in length
48882	48981	gap of 100 bp
48982	50240	config of 1255 bp in length
50241	50340	gap of 100 bp
50341	51544	config of 1204 bp in length
51545	51644	gap of 100 bp
51645	52828	config of 1184 bp in length
52829	52928	gap of 100 bp
52929	54158	config of 1222 bp in length
54158	54257	gap of 100 bp
54258	54450	config of 1199 bp in length
54451	55550	gap of 100 bp
55551	56731	config of 1181 bp in length
56732	56831	gap of 100 bp
56832	58019	config of 1188 bp in length
58020	58119	gap of 100 bp
58120	59285	config of 1166 bp in length
59286	59385	gap of 100 bp
59386	60525	config of 1140 bp in length
60526	60625	gap of 100 bp
60626	61717	config of 1092 bp in length
61718	61817	gap of 100 bp
61818	62991	config of 1174 bp in length
62992	63091	gap of 100 bp
63092	64267	config of 1176 bp in length
64268	64367	gap of 100 bp
64368	65552	config of 1185 bp in length
65553	65652	gap of 100 bp
65653	66853	config of 1201 bp in length
66854	66953	gap of 100 bp
66954	68145	config of 1192 bp in length
68146	68245	gap of 100 bp
68246	69423	config of 1176 bp in length
69424	69523	gap of 100 bp
69524	70737	config of 1214 bp in length
70738	70837	gap of 100 bp
70838	71997	config of 1160 bp in length
71998	72097	gap of 100 bp
72098	73254	config of 1157 bp in length

RESULT 9	AC118192/c	AC118192	187353 bp	DNA	linear	HTG 14-APR-2007
LOCUS		Mus musculus clone RP23-200C17,				
DEFINITION		pieces.				
ACCESSION		AC118192				
VERSION		AC118192.1	GI:20147957			
KEYWORDS		HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE		Mus musculus (house mouse)				
ORGANISM		Mus musculus				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS		1 (bases 1 to 187353)				
TITLE		Birten,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL		Mus musculus, clone RP23-200C17				
REFERENCE		unpublished				
AUTHORS		2 (bases 1 to 187353)				
		Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,				

TITLE	Direct Submission
JOURNAL	Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker:

----- Genome Center  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----

Center project name: L19568  
Center clone name: 200\_C\_17

----- Summary Statistics -----

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
Consensus quality: 180258 bases at least 400

Consensus quality: 184259 bases at least Q40  
Consensus quality: 185710 bases at least Q30  
Consensus quality: 186168 bases at least Q20

Insert size: 194000 : agarose-  
consensus quality: 186189 bases at 1eabc qzv

Insert size: 186453; sum-of-contigs: 186453; quality coverage: 7.4 in 920 bases.

Quality coverage: 7.7 in 020 bases; sum-of-contigs

[illegible]

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 10 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* Of the gaps between them are based on estimates that in

\* This sequence will be replaced  
provided by the submitter:

\* by the finished sequence as soon as it is available and



VERSION AX282592.1 GI:16609675  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Iesakant, S.D., Huang, J., Sheung, J. and Pray, T.R.  
TITLE Ubiquitin ligase assay  
JOURNAL Patent: WO 0175145-A 7 11-OCT-2001;  
Rigel Pharmaceutical, Inc. (US)  
FEATURES  
source 1..342  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 82 a 84 c 108 g 68 t  
ORIGIN

Query Match 25.2%; Score 32; DB 6; Length 342;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45  
151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 12  
AF142060 342 bp mRNA linear PRI 14-JUL-1999  
LOCUS Homo sapiens RING finger protein (ROC2) mRNA, complete cds.  
DEFINITION AF142060  
ACCESSION AF142060  
VERSION AF142060.1 GI:4809217  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 342)  
AUTHORS Ohta, T., Michel, J.J., Schotelius, A.J. and Xiong, Y.  
TITLE ROC2, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity  
JOURNAL Mol. Cell 3 (4), 535-541 (1999)  
MEDLINE 99247022  
PUBMED 10230407  
REFERENCE 2 (bases 1 to 342)  
AUTHORS Ohta, T., Michel, J.J. and Xiong, Y.  
TITLE Direct Submission  
JOURNAL Submitted (10-APR-1999) Linberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Mason Farm Rd. and Manning Dr., Chapel Hill, NC 27599-7295, USA  
FEATURES  
source 1..342  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..342  
/gene="ROC2"  
1..342  
/gene="ROC2"  
/note="Cullin partnering protein with associated ubiquitin ligase activity"  
/codon\_start=1  
/product="RING finger protein"  
/protein\_id="AAD30147.1"  
/db\_xref="GI:4809218"  
/translation="MADVDEGETCALASHSSGSGSGDKMPSLKKNAYAMWSMD  
VECDICAIQRYQVMDACLRCAQENKQEDCVVWGBSCNHSFNHCNSLWVKQNNRCPLC  
QODWVQRIQK"  
BASE COUNT 82 a 83 c 109 g 68 t  
ORIGIN

Query Match 25.2%; Score 32; DB 9; Length 342;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45  
151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 13  
BT007348 342 bp mRNA linear PRI 13-MAY-2003  
LOCUS Homo sapiens ring finger protein 7 mRNA, complete cds.  
DEFINITION BT007348  
ACCESSION BT007348.1 GI:30583534  
VERSION BT007348.1 GI:30583534  
KEYWORDS FLI-CDNA.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 342)  
AUTHORS Kainline, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y., Pheasant, M. and Farmer, A.  
TITLE Cloning of human full-length CDSs in BD Creator(TM) system donor vector  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 342)  
Kainline, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y., Pheasant, M. and Farmer, A.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA  
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the Sali and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after Sali site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orclones>.  
FEATURES  
source 1..342  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="GH00912X1.0"  
/clone\_lib="BD Creator(TM) CDS Library derived from MGC collection"  
/lab\_host="DH5alpha T1 resistant"  
/note="Vector: pDNR-Dual"  
1..342  
/codon\_start=1  
/product="ring finger protein 7"  
/protein\_id="AAP36012.1"  
/db\_xref="GI:30583535"  
/translation="MADVDEGETCALASHSSGSGSGDKMPSLKKNAYAMWSMD  
VECDICAIQRYQVMDACLRCAQENKQEDCVVWGBSCNHSFNHCNSLWVKQNNRCPLC  
QODWVQRIQK"  
BASE COUNT 82 a 83 c 109 g 68 t  
ORIGIN

Query Match 25.2%; Score 32; DB 9; Length 342;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45  
151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 14  
BD096967 754 bp DNA linear PAT 27-AUG-2002  
LOCUS BD096967 SAG:apoptosis sensitivity gene.  
DEFINITION BD096967  
ACCESSION BD096967.1 GI:22642555  
VERSION JP 2001526063-A/2.  
KEYWORDS  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 2 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/2  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC /note = 'Human SAG'  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT misc\_feature 1..754.  
Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 205 a 155 c 201 g 193 t  
ORIGIN  
Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 14 GCCATCTGCAGGGTCCAGGTGATGATGCTG 45  
151 GCCATCTGCAGGGTCCAGGTGATGATGCTG 182  
Db  
RESULT 15  
BD096981 754 bp DNA linear PAT 27-AUG-2002  
LOCUS BD096981 SAG:apoptosis sensitivity gene.  
DEFINITION BD096981  
ACCESSION BD096981.1 GI:22642569  
VERSION JP 2001526063-A/16.  
KEYWORDS  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 16 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/16  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC /note = 'Human SAG'  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT misc\_feature 1..754.  
Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 206 a 155 c 201 g 192 t  
ORIGIN  
Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 14 GCCATCTGCAGGGTCCAGGTGATGATGCTG 45  
151 GCCATCTGCAGGGTCCAGGTGATGATGCTG 182  
Db

PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 206 a 155 c 201 g 192 t  
ORIGIN  
Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 14 GCCATCTGCAGGGTCCAGGTGATGATGCTG 45  
151 GCCATCTGCAGGGTCCAGGTGATGATGCTG 182  
Db  
RESULT 16  
BD096985 754 bp DNA linear PAT 27-AUG-2002  
LOCUS BD096985 SAG:apoptosis sensitivity gene.  
DEFINITION BD096985  
ACCESSION BD096985.1 GI:22642573  
VERSION JP 2001526063-A/20.  
KEYWORDS  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 20 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/20  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 206 a 155 c 201 g 192 t  
ORIGIN  
Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 14 GCCATCTGCAGGGTCCAGGTGATGATGCTG 45  
151 GCCATCTGCAGGGTCCAGGTGATGATGCTG 182  
Db



Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 17  
LOCUS BD096986 754 bp DNA linear PAT 27-AUG-2002  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096986.1 GI:22642574  
VERSION BD096986.1  
KEYWORDS JP 2001526063-A/21.  
SOURCE unclassified  
ORGANISM unclassified

REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 21 18-DEC-2001;  
WARNER LAMBERT CO

COMMENT  
OS Unidentified  
PN JP 2001526063-A/21  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat peptide 1..339  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 207 a 154 c 201 g 192 t

ORIGIN

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45  
|||||  
151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 18  
LOCUS BD096987 754 bp DNA linear PAT 27-AUG-2002  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096987.1 GI:22642575  
VERSION BD096987.1  
KEYWORDS JP 2001526063-A/22.  
SOURCE unclassified  
ORGANISM unclassified

REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 22 18-DEC-2001;  
WARNER LAMBERT CO

COMMENT  
OS Unidentified  
PN JP 2001526063-A/22  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat peptide 1..339  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 153 c 202 g 193 t

ORIGIN

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45  
|||||  
151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

Db 151 GCCATCTGCAGGTCACAGTGATGATGCCTG 182

RESULT 19  
LOCUS BD096988 754 bp DNA linear PAT 27-AUG-2002  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096988.1 GI:22642576  
VERSION BD096988.1  
KEYWORDS JP 2001526063-A/23.  
SOURCE unclassified  
ORGANISM unclassified

REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 23 18-DEC-2001;  
WARNER LAMBERT CO

COMMENT  
OS Unidentified  
PN JP 2001526063-A/23  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat peptide 1..339  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t

ORIGIN

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GCCATCTGCAGGTCACAGTGATGATGCCTG 45

```

Db      151 GCCATCTGCAGGCTCCAGGTGATGATGCTG 182
|||||
RESULT 20
LOCUS   BD096989                      754 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION
SAG:apoptosis sensitivity gene.
ACCESSION   BD096989
VERSION     BD096989.1 GI:22642577
KEYWORDS    JP 2001526063-A/24.
SOURCE      unidentified
ORGANISM    unidentified
REFERENCE   1 (bases 1 to 754)
AUTHORS     Sun,Y.
TITLE       SAG:apoptosis sensitivity gene
JOURNAL     Patent: JP 2001526063-A 24 18-DEC-2001;
            WARNER LAMBERT CO
COMMENT     OS Unidentified
            PN JP 2001526063-A/24
            PD 18-DEC-2001
            PF 15-DEC-1998 JP 2000525451
            PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI
            YI SUN

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      14 GCCATCTGCAGGCTCCAGGTGATGATGCTG 45
|||||
Db      151 GCCATCTGCAGGCTCCAGGTGATGATGCTG 182
|||||

FEATURES
source
1..754
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT      206 a 155 c 201 g 192 t

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      14 GCCATCTGCAGGCTCCAGGTGATGATGCTG 45
|||||
Db      151 GCCATCTGCAGGCTCCAGGTGATGATGCTG 182
|||||

RESULT 21
LOCUS   BD096990                      754 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION
SAG:apoptosis sensitivity gene.
ACCESSION   BD096990
VERSION     BD096990.1 GI:22642578
KEYWORDS    JP 2001526063-A/25.
SOURCE      unidentified
ORGANISM    unidentified
REFERENCE   1 (bases 1 to 754)
AUTHORS     Sun,Y.
TITLE       SAG:apoptosis sensitivity gene
JOURNAL     Patent: JP 2001526063-A 25 18-DEC-2001;
            WARNER LAMBERT CO
COMMENT     OS Unidentified
            PN JP 2001526063-A/25
            PD 18-DEC-2001
            PF 15-DEC-1998 JP 2000525451
            PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI
            YI SUN

```

```

YI SUN
PC      C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,
PC      A61P39/06,
PC      A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC
,C12N5/10,C12Q1/68,
PC      G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC
Strandedness: Double;
CC      Topology: linear;
CC      SAG:apoptosis sensitivity gene
FH      Key      Location/Qualifiers
FT      CDS      1..339
FT      mat_peptide 1..339.
FT      Location/Qualifiers
source
1..754
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT      206 a 155 c 201 g 192 t

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      14 GCCATCTGCAGGCTCCAGGTGATGATGCTG 45
|||||
Db      151 GCCATCTGCAGGCTCCAGGTGATGATGCTG 182
|||||

RESULT 22
LOCUS   BD096991                      754 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION
SAG:apoptosis sensitivity gene.
ACCESSION   BD096991
VERSION     BD096991.1 GI:22642579
KEYWORDS    JP 2001526063-A/26.
SOURCE      unidentified
ORGANISM    unidentified
REFERENCE   1 (bases 1 to 754)
AUTHORS     Sun,Y.
TITLE       SAG:apoptosis sensitivity gene
JOURNAL     Patent: JP 2001526063-A 26 18-DEC-2001;
            WARNER LAMBERT CO
COMMENT     OS Unidentified
            PN JP 2001526063-A/26
            PD 18-DEC-2001
            PF 15-DEC-1998 JP 2000525451
            PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI
            YI SUN

PC      C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,
PC      A61P39/06,
PC      A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC
,C12N5/10,C12Q1/68,
PC      G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC
Strandedness: Double;
CC      Topology: linear;
CC      SAG:apoptosis sensitivity gene
FH      Key      Location/Qualifiers
FT      CDS      1..339
FT      mat_peptide 1..339.
FT      Location/Qualifiers
source
1..754
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT      206 a 155 c 201 g 192 t

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      14 GCCATCTGCAGGCTCCAGGTGATGATGCTG 45
|||||
Db      151 GCCATCTGCAGGCTCCAGGTGATGATGCTG 182
|||||

```

QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45  
Db 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182

RESULT 23  
BD096992 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096992  
ACCESSION BD096992  
VERSION BD096992.1 GI:22642580  
KEYWORDS JP 2001526063-A/27.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 27 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/27  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FT Location/Qualifiers  
1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45  
Db 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182

RESULT 24  
BD096994 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096994  
ACCESSION BD096994  
VERSION BD096994.1 GI:22642582  
KEYWORDS JP 2001526063-A/29.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 29 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/29  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451

PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FT Location/Qualifiers  
1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 207 a 155 c 201 g 191 t

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45  
Db 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182

RESULT 25  
BD096995 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096995  
ACCESSION BD096995  
VERSION BD096995.1 GI:22642583  
KEYWORDS JP 2001526063-A/30.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 30 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/30  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FT Location/Qualifiers  
1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t

Query Match 25.2%; Score 32; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCACAGTGTGATGCTG 45  
 DB 151 GCCATCTGCAGGTCACAGTGTGATGCTG 182

RESULT 26  
 AF092878  
 LOCUS AF092878 754 bp mRNA linear PRI 24-JUL-2001  
 DEFINITION Homo sapiens zinc RING finger protein SAG mRNA, complete cds.  
 ACCESSION AF092878  
 VERSION AF092878.1 GI:4588033  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 754)  
 Tian, Y., Wang, Y., Aviram, M., Swaroop, M., Loo, J.A., Bian, J.,  
 SAG, a novel zinc RING finger protein that protects cells from  
 apoptosis induced by redox agents  
 Mol. Cell. Biol. 19 (4), 3145-3155 (1999)  
 99182502  
 10082581  
 2 (bases 1 to 754)  
 Sun, Y.  
 Alterations of SAG mRNA in human cancer cell lines: requirement for  
 the RING finger domain for apoptosis protection  
 Carcinogenesis 20 (10), 1899-1903 (1999)  
 99435944  
 10506102  
 3 (bases 1 to 754)  
 Swaroop, M., Bian, J., Aviram, M., Duan, H., Bisgaier, C.L., Loo, J.A.  
 and Sun, Y.  
 Expression, purification, and biochemical characterization of SAG,  
 a RING finger redox-sensitive protein  
 Free Radical Biol. Med. 27, 193-202 (1999)  
 4 (bases 1 to 754)  
 Swaroop, M., Wang, Y., Miller, P., Duan, H., Jackoe, T., Madore, S.J. and  
 Sun, Y.  
 Yeast homolog of human SAG/ROC2/Rbx2/Hrt2 is essential for cell  
 growth, but not for germination: chip profiling implicates its role  
 in cell cycle regulation  
 Oncogene 19 (24), 2855-2866 (2000)  
 20309864  
 10851089  
 5 (bases 1 to 754)  
 Duan, H., Tsvetkov, L.M., Liu, Y., Song, Y., Swaroop, M., Wen, R.,  
 Kung, H.F., Zhang, H. and Sun, Y.  
 Promotion of S-phase entry and cell growth under serum starvation  
 by SAG/ROC2/Rbx2/Hrt2, an E3 ubiquitin ligase component:  
 association with inhibition of p27 accumulation  
 Mol. Carcinog. 30 (1), 37-46 (2001)  
 21152847  
 11255262  
 6 (bases 1 to 754)  
 Sun, Y.  
 Direct Submision  
 Submitted (16-SEP-1998) Department of Molecular Biology,  
 Parke-Davis, 2800 Plymouth Rd, Ann Arbor, MI 48105, USA

FEATURES  
 SOURCE  
 1..754  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3q22-q24"  
 /cell\_line="Hela, D98/AH-2, HPRT-"  
 1..372  
 /function="growth promotion"  
 /note="redox sensitive, metal binding; expression protects  
 cells from apoptosis induced by redox compounds"

CDS

BASE COUNT 205 a 155 c 201 g 193 t

ORIGIN

Query Match  
 Best Local Similarity 100.0%; Pred. No. 4e-08;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCACAGTGTGATGCTG 45  
 DB 151 GCCATCTGCAGGTCACAGTGTGATGCTG 182

RESULT 27  
 BC008627  
 LOCUS BC008627 816 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, ring finger protein 7, clone MGC:17274 IMAGE:4177613,  
 mRNA, complete cds.  
 ACCESSION BC008627  
 VERSION BC008627.1 GI:14250388  
 KEYWORDS  
 SOURCE MGC.  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 816)  
 Strausberg, R.  
 Direct Submission  
 Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 COMMENT  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: David N. Louis, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: villalob@bcm.tmc.edu.  
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov  
 Series: IRAK Plate: 12 Row: n Column: 14  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

FEATURES  
 SOURCE  
 1..816  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocustID:9616"  
 /db\_xref="taxon:9606"  
 /clone="MGC:17274 IMAGE:4177613"  
 /tissue type="Brain, anaplastic oligodendroglioma with  
 1p/19q loss"  
 /clone\_id="NCI CGAP\_Brn67"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 37..378  
 /codon\_start=1

CDS

```

/product="ring finger protein 7"
/protein_id="AA08627.1"
/db_xref="GI:14250389"
/translation="MADVDGEETCALASHSSGSGSGDKMFLKKNAVAMWSMD
VECDTCAICRVQMDACLRQAEKQEDCVVWVGECNHSFHNCSLWVKONRRCPCLC
QODWVVRIGK"
BASE COUNT      226 a      178 c      213 g      199 t
ORIGIN
Query Match      25.2%; Score 32; DB 9; Length 816;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      14 GCCATCTGCAGGTCGAGTGATGATGCCTG 45
Db      187 GCCATCTGCAGGTCGAGTGATGATGCCTG 218

RESULT 28
LOCUS      BC005966      822 bp      mRNA      linear      PRI 12-JUL-2001
DEFINITION Homo sapiens, ring finger protein 7, clone MGC:14618 IMAGE:4069078,
            mRNA, complete cds.
ACCESSION      BC005966
VERSION      BC005966.1 GI:13543635
KEYWORDS      MGC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 822)
AUTHORS      Strausberg, R.
TITLE      Direct Submission
JOURNAL      Submitted (02-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
            Email: cgabs-rc@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: 0 Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene
prediction, Similarity but not identity to protein.
FEATURES
source
1..822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:9616"
/db_xref="taxon:9606"
/clone="MGC:14618 IMAGE:4069078"
/tissue_type="Kidney, hypernephroma"
/clone_lib="NIH_MGC_58"
/lab_host="DH10B"
/notes="Vector: PDNR-LIB"
27..368
/codon_start=1
/product="ring finger protein 7"
/protein_id="AA08627.1"
/db_xref="GI:13543636"
/translation="MADVDGEETCALASHSSGSGSGDKMFLKKNAVAMWSMD

```

```

VECDTCAICRVQMDACLRQAEKQEDCVVWVGECNHSFHNCSLWVKONRRCPCLC
QODWVVRIGK"
BASE COUNT      240 a      175 c      210 g      197 t
ORIGIN
Query Match      25.2%; Score 32; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      14 GCCATCTGCAGGTCGAGTGATGATGCCTG 45
Db      177 GCCATCTGCAGGTCGAGTGATGATGCCTG 208

RESULT 29
LOCUS      AF164679      836 bp      mRNA      linear      PRI 14-OCT-1999
DEFINITION Homo sapiens ring finger protein CKBP1 mRNA, complete cds.
ACCESSION      AF164679
VERSION      AF164679.1 GI:5917673
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 836)
AUTHORS      Son, M.-Y., Park, J.-W., Kim, Y.-S., Kang, S.-W., Marshak, D.R., Park, W. and
            Bae, Y.-S.
TITLE      Protein kinase CKII interacts with and phosphorylates the SAG
            protein containing ring-H2 finger motif
JOURNAL      Biochem. Biophys. Res. Commun. 263 (3), 743-748 (1999)
MEDLINE      99443734
PubMed      10512750
REFERENCE      2 (bases 1 to 836)
AUTHORS      Son, M.-Y., Park, J.-W., Kim, Y.-S., Kang, S.-W., Marshak, D.R., Park, W.
            and Bae, Y.-S.
TITLE      Direct Submission
JOURNAL      Submitted (01-JUL-1999) Department of Biochemistry, Kyungpook
            National University, Taegu 702-701, Korea
FEATURES
source
1..836
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
54..395
/notes="SAG, contains an H2 ring finger motif"
/codon_start=1
/product="ring finger protein CKBP1"
/protein_id="AAD55984.1"
/db_xref="GI:5917674"
/translation="MADVDGEETCALASHSSGSGSGDKMFLKKNAVAMWSMD
VECDTCAICRVQMDACLRQAEKQEDCVVWVGECNHSFHNCSLWVKONRRCPCLC
QODWVVRIGK"
BASE COUNT      228 a      185 c      223 g      200 t
ORIGIN
Query Match      25.2%; Score 32; DB 9; Length 836;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      14 GCCATCTGCAGGTCGAGTGATGATGCCTG 45
Db      204 GCCATCTGCAGGTCGAGTGATGATGCCTG 235

RESULT 30
LOCUS      AC118192      187353 bp      DNA      linear      HTG 14-APR-2002
DEFINITION Mus musculus clone RP23-200C17, WORKING DRAFT SEQUENCE, 10 ordered
            pieces.
ACCESSION      AC118192
VERSION      AC118192.1 GI:20147957
KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

```

SOURCE	ORGANISM
Mus musculus	(house mouse)
Eukaryote;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia;	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE	
1 (bases 1 to 187353)	
Bioren,B., Linton,L., Nussbaum,C. and Lander,E.	
Mus musculus, clone RP23-200C17	
Unpublished	
2 (bases 1 to 187353)	
Bioren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,	
Anderson,S., Barina,N., Bastien,E., Bloom,T., Boguslavsky,L.,	
Buckshalter,B., Brown,A., Camarata,J., Campionio,A., Chang,J.,	
Chazaro,B., Choepel,Y., Colangelo,K., Collins,S., Collymore,A.,	
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,	
Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gairyna,S	
Grinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,	
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,	
Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,	
Landers,T., Lehoczek,J., Levine,R., Lindblad-Toh,K., Liu,G.,	
Maclean,C., MacDonald,P., Major,J., Margulis,N., Mathews,C.,	
McCarthy,M., McSwan,P., McKernan,K., Meltrin,J., Neneus,L.,	
Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,	
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,	
Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,	
Ramond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,	
Roman,J., Rossetti,M., Roy,A., Santos,R., Schauer,S., Schnepack,R.	
Seeman,S., Severy,P., Spencer,B., Stange-Rothman,N., Stojanovic,N	
Strasus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,	
Tophan,K., Travers,M., Travis,N., Trifillo,D., Vassiliev,H.,	
Viel,R., Vo,A., Wilson,B., Wu,X., Wymen,D., Ye,W.T., Young,G.,	
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	
Direct Submission	
Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome	
Research, 320 Charles Street, Cambridge, MA 02141, USA	
All repeats were identified using RepeatMasker:	
Smt, A.F.A. & Green, P. (1996-1997)	
http://ftp.genome.washington.edu/RM/RepeatMasker.html	
Genome Center	
Center: Whitehead Institute/ MIT Center for Genome Research	
Center code: WIBR	
Web site: http://www-seq.wi.mit.edu	
Contact: sequence.submissions@genome.wi.mit.edu	
Project Information	
Center project name: L19568	
Center clone name: 200 C-17	
Summary Statistics	
Sequencing vector: plasmid; n/a; 100% of reads	
Chemistry: Dye-terminator Big Dye; 100% of reads	
Assembly program: Phrap; version 0.960731	
Consensus quality: 184259 bases at least Q40	
Consensus quality: 185710 bases at least Q30	
Consensus quality: 186169 bases at least Q20	
Insert size: 194000; agarose-fp	
Insert size: 186453; sum-of-ctrls	
Quality coverage: 7.4 in Q20 bases; agarose-fp	
Quality coverage: 7.7 in Q20 bases; sum-of-ctrls	
NOTE: This is a 'working draft' sequence. It currently	
consists of 10 contigs. Gaps between the contigs	
are represented as runs of N. The order of the pieces	
is believed to be correct as given, however, the sizes	
of the gaps between them are based on estimates that have	
provided by the submittor.	
This sequence will be replaced	
by the finished sequence as soon as it is available and	
the accession number will be preserved.	
1	629: contig of 629 bp in length
630	729: gap of 100 bp
730	750: contig of 21 bp in length
751	850: gap of 100 bp
851	3312: contig of 2362 bp in length
3213	3312: gap of 100 bp
3313	6908: contig of 3596 bp in length

FEATURES	SOURCE
* 6909	7008: gap of 100 bp
* 7009	14434: contig of 7426 bp in length
* 14435	14534: gap of 100 bp
* 14535	32181: contig of 17647 bp in length
* 32182	32281: gap of 100 bp
* 32282	51167: contig of 18886 bp in length
* 51168	51267: gap of 100 bp
* 51268	72870: contig of 21603 bp in length
* 72871	72970: gap of 100 bp
* 149992	149992: contig of 77022 bp in length
* 149993	150092: gap of 100 bp
* 150093	187353: contig of 37261 bp in length.
Location/Qualifiers	
1. 187353	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/clone="RP23-200C17"	
/clone_1fb="RP23 Female Mouse BAC"	
1. 629	
/note="assembly_fragment"	
730. 750	
/note="assembly_fragment"	
851. 3212	
/note="assembly_fragment"	
3313. 6908	
/note="assembly_fragment"	
7009. 14434	
/note="assembly_fragment"	
14535. 32181	
/note="assembly_fragment"	
32282. 51167	
/note="assembly_fragment"	
51268. 72870	
/note="assembly_fragment"	
72971. 149992	
/note="assembly_fragment"	
150093. 187353	
/note="assembly_fragment"	
clone_end:T7	
vector_side:right"	
BASE COUNT	60628 a 34855 c 36129 g 54836 t 905 others
ORIGIN	
Query Match	25.2%; Score 32; DB 2; Length 187353;
Best Local Similarity	100.0%; Prid. No. 3.6e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Py	13 TGGCATCTGCAGGGTCCAGGTGATGATGCT 44
Db	46892 TGGCATCTGCAGGGTCCAGGTGATGATGCT 46923
RESULT 31	
AC095698/c	
LOCUS	AC095698 254832 bp DNA linear HTG 09-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-9D13, WORKING DRAFT SEQUENCE.
ACCESSION	AC095698
KEYWORDS	AC095698.6 GI:24817898
VERSIONS	HTG: HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
	1 (bases 1 to 254832)
	Muzny,D.,Marle, Metzker,M.,Lee, Abramzon,S., Adams,C., Alder,J.,
	Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
	Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
	Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
	Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
	Bryant,N., Buay,C., Burch,P., Burrill,K., Calderon,E.,
	Cadenas,V., Carter,K., Cavazos,I., Cessari,H., Center,A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Detamo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunatirige, P., Haaland, W., Hamill, C., Hamilton, C., Hernandez, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladik, S. L., Hodgson, A., Hogue, M., Hollins, B., Howell, S., Hu, Y., Hume, J., Idelberg, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Kowals, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensueta, L., Louie, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedim, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polinder, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, U., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Sherry, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soos, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, U., Zhou, J., Zhou, J., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Unpublished  
Direct Submission  
2 (bases 1 to 254832)  
Worley, K. C.  
Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 254832)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 9, 2002 this sequence version replaced gi:23267899.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GDBC  
Center clone name: CH230-9D13  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 238918 bases at least Q40  
Consensus quality: 241069 bases at least Q30  
Consensus quality: 242304 bases at least Q20  
Estimated insert size: 243228; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 254832: contig of 254832 bp in length.  
Location/Qualifiers  
1. 254832  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-9D13"  
1. 1482  
/note="wgs end extension  
clone\_end:5p6"  
2092. 2921  
/note="clone\_boundary  
clone\_end:5p6  
site:EcORI  
end sequence:BH339916"  
23357. 235284  
/note="wgs contig"  
253718. 254832  
/note="wgs contig"  
complement(253801..254228)  
/note="clone\_boundary  
clone\_end:T7  
site:EcORI  
end sequence:BH339912"  
BASE COUNT 63080 a 58350 c 57410 g 64047 t 11945 others  
ORIGIN  
Query Match 24.4%; Score 31; DB 2; Length 254832;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 CCATCGCAGGGTCCAGGATGATGCTG 45  
Db 146354 CCATCGCAGGGTCCAGGATGATGCTG 146324  
-----  
RESULT 32  
BD096984  
LOCUS BD096984 754 bp DNA linear PAT 27-AUG-2002  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096984.1 GI:22642572  
VERSION BD096984.1 GI:22642572  
KEYWORDS JP 2001526063-A/19.  
SOURCE unidentified  
ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 19 18-DEC-2001;

COMMENT WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/19  
PD 18-DEC-2001  
PR 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT Location/Qualifiers  
1..754  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t

ORIGIN

Query Match 23.6%; Score 30; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 6.9e-07;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCACGATGATGATGCC 43  
Db 151 GCCATCTGCAGGTCACGATGATGATGCC 180

RESULT 33  
BD096993 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096993.1 GI:22642581  
VERSION BD096993.1 GI:22642581  
KEYWORDS JP 2001526063-A/28.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun,Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 28 18-DEC-2001;  
WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/28  
PD 18-DEC-2001  
PR 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT Location/Qualifiers  
1..754  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 207 a 155 c 201 g 191 t

ORIGIN

Query Match 23.6%; Score 30; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 6.9e-07;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCACGATGATGATGCC 43  
Db 151 GCCATCTGCAGGTCACGATGATGATGCC 180

RESULT 34  
AC119145 148290 bp DNA linear HTG 11-JUN-2002  
LOCUS Rattus norvegicus chromosome 1 clone RP32-4G14, WORKING DRAFT  
DEFINITION SEQUENCE, 10 unordered pieces.  
AC119145  
AC119145.1 GI:20303534  
HTG, HTGS\_PHASE1, HTGS\_FULFILL, HTGS\_DRAFT.  
VERSION Rattus norvegicus (Norway rat)  
KEYWORDS Rattus norvegicus  
SOURCE Rattus norvegicus  
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 148290)  
Taenzer,S., Monti,U., Gloeckner,G., Goesele,C., Baumgart,C.,  
Huebner,N. and Platzner,M.  
Rat chromosome 1 genomic sequence  
Unpublished  
2 (bases 1 to 148290)  
Taenzer,S. and Platzner,M.  
Direct Submission  
Submitted (25-APR-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany  
3 (bases 1 to 148290)  
Taenzer,S. and Platzner,M.  
Direct Submission  
Submitted (11-JUN-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany  
----- Genome Center  
Center: Institute of Molecular Biotechnology  
Center code: IMB  
Web site: http://genome.imb-jena.de/  
Contact: gscj-submit@genome.imb-jena.de  
----- Project Information  
Center project name: RP32-4G14  
Center clone name: RP32-4G14  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 140730 bases at least Q40  
Consensus quality: 143676 bases at least Q30  
Consensus quality: 145595 bases at least Q20  
Quality coverage: 8.49

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will



RESULT 35	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AC125667	AC125667	Rattus norvegicus clone CH230-21C17, WORKING DRAFT SEQUENCE, 2	AC125667	AC125667	HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	1 (bases 1 to 250425)	Murphy, D. Marie, Metzker, M. Lee, Adrianzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguitano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bernhamed, F., Biwalto, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, J., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Cemer, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Benson, S., Derasmo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Friaser, C. M., Gabisi, A., Gall, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, B., Geer, K., Giller, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hildun, S. L., Hodgson, A., Hognes, M.,

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Jackson, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
 Hollins, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolliffe, A.,  
 Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshaw, L., Loulesged, H., Lozano, R. J., Lu, X., Ma, J.,  
 Meshawati, M., Mahindaratne, M., Mahmud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
 Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Miodovlevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Mookeleneh, O., Okunou, G., Olarnpusoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poldexter, A., Popovic, D., Plimus, E., Pu, L.,  
 Piazzi, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shamsan, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sotelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasa, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Wilson, R., Wlezyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausen, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G., and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 250425).  
 Worley, K.C.  
 Direct Submission  
 Submitted (29-JUN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 250425)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:24819649.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GGLK  
 Center clone name: CH230-21C17  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0,  
 Consensus quality: 241733 bases at least Q40  
 Consensus quality: 244050 bases at least Q30  
 Consensus quality: 245370 bases at least Q20  
 Estimated insert size: 249556; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation



```

BASE COUNT      204 a 155 c 197 g 191 t
ORIGIN
Query Match      22.8%; Score 29; DB 6; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGCTGGCGGAGA 99
DB 201 CAAGAGACTGTGTGTGCTGGCGGAGA 229

RESULT 38
BD096982      754 bp      DNA      linear      PAT 27-AUG-2002
LOCUS
DEFINITION    SAG:apoptosis sensitivity gene.
ACCESSION    BD096982
VERSION      BD096982.1 GI:22642570
KEYWORDS      JP 2001526063-A/17.
SOURCE      unidentified
ORGANISM      unidentified
REFERENCE      1 (bases 1 to 754)
AUTHORS      Sun, Y.
TITLE      SAG:apoptosis sensitivity gene
JOURNAL      Patent: JP 2001526063-A 17 18-DEC-2001;
              WARNER LAMBERT CO
COMMENT      OS Unidentified
              PN JP 2001526063-A/17
              PD 18-DEC-2001
              PE 15-DEC-1998 JP 2000525451
              PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI
              YI SUN
              PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,
              PC A61P39/06,
              PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC
              , C12N5/10, C1201/68,
              PC G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC
              Strandedness: Double;
              CC Topology: Linear;
              CC SAG:apoptosis sensitivity gene
              FH Key      Location/Qualifiers
              FT CDS      1..339
              FT mat_peptide 1..339.
              FT Location/Qualifiers
              1..754
              /organism="unidentified"
              /mol_type="genomic DNA"
              /db_xref="taxon:32644"

BASE COUNT      206 a 155 c 201 g 192 t
ORIGIN
Query Match      22.8%; Score 29; DB 6; Length 754;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGCTGGCGGAGA 99
DB 208 CAAGAGACTGTGTGTGCTGGCGGAGA 236

RESULT 39
BD096983      754 bp      DNA      linear      PAT 27-AUG-2002
LOCUS
DEFINITION    SAG:apoptosis sensitivity gene.
ACCESSION    BD096983
VERSION      BD096983.1 GI:22642571
KEYWORDS      JP 2001526063-A/18.
SOURCE      unidentified
ORGANISM      unidentified
REFERENCE      1 (bases 1 to 754)

```

```

AUTHORS      Sun, Y.
TITLE      SAG:apoptosis sensitivity gene
JOURNAL      Patent: JP 2001526063-A 18 18-DEC-2001;
              WARNER LAMBERT CO
COMMENT      OS Unidentified
              PN JP 2001526063-A/18
              PD 18-DEC-2001
              PE 15-DEC-1998 JP 2000525451
              PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI
              YI SUN
              PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,
              PC A61P39/06,
              PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC
              , C12N5/10, C1201/68,
              PC G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC
              Strandedness: Double;
              CC Topology: Linear;
              CC SAG:apoptosis sensitivity gene
              FH Key      Location/Qualifiers
              FT CDS      1..339
              FT mat_peptide 1..339.
              FT Location/Qualifiers
              1..754
              /organism="unidentified"
              /mol_type="genomic DNA"
              /db_xref="taxon:32644"

BASE COUNT      207 a 155 c 201 g 191 t
ORIGIN
Query Match      22.8%; Score 29; DB 6; Length 754;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGCTGGCGGAGA 99
DB 208 CAAGAGACTGTGTGTGCTGGCGGAGA 236

RESULT 40
AC103666
LOCUS
DEFINITION    Mus musculus clone RP23-397H13, LOW-PASS SEQUENCE SAMPLING.
ACCESSION    AC103666
VERSION      AC103666.2 GI:27311471
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 112321)
Birren, B., Nusbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 112321)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barta, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, D., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneu, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C.,
North, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

```

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 112321)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barns, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagos, B., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C.,  
Kamat, A., Karakas, A., Kellis, C., Landers, T., Levine, R.,  
Lindblad-Ton, K., Liu, G., MacLean, C., MacDonald, P., Major, J.,  
Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Piere, N., Raymond, C., Retta, R.,  
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnuppach, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K.,  
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 20, 2002 this sequence version replaced gi:17149550.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L19615  
Center clone name: 397\_H\_13

\* NOTE: This record contains 89 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 1147: contig of 1147 bp in length  
\* 1148 1247: gap of 100 bp  
\* 1248 2348: contig of 1101 bp in length  
\* 2349 2448: gap of 100 bp  
\* 2449 3511: contig of 1063 bp in length  
\* 3512 3611: gap of 100 bp  
\* 3612 4746: contig of 1135 bp in length  
\* 4747 4846: gap of 100 bp  
\* 4847 6030: contig of 1184 bp in length  
\* 6031 6130: gap of 100 bp  
\* 6131 7290: contig of 1160 bp in length  
\* 7291 7390: gap of 100 bp  
\* 7391 8577: contig of 1187 bp in length  
\* 8578 8678: gap of 100 bp  
\* 8679 9937: contig of 1260 bp in length  
\* 9938 10037: gap of 100 bp  
\* 10038 11192: contig of 1155 bp in length  
\* 11193 11292: gap of 100 bp  
\* 11293 12454: contig of 1162 bp in length  
\* 12455 12554: gap of 100 bp  
\* 12555 13718: contig of 1164 bp in length

13719 13818: gap of 100 bp  
\* 13819 15002: contig of 1184 bp in length  
\* 15003 15102: gap of 100 bp  
\* 15103 16234: contig of 1132 bp in length  
\* 16235 16334: gap of 100 bp  
\* 16335 17461: contig of 1127 bp in length  
\* 17462 17561: gap of 100 bp  
\* 17562 18650: contig of 1089 bp in length  
\* 18651 18750: gap of 100 bp  
\* 18751 19912: contig of 1162 bp in length  
\* 19913 20013: gap of 100 bp  
\* 20013 21199: contig of 1187 bp in length  
\* 21200 21299: gap of 100 bp  
\* 21300 22462: contig of 1163 bp in length  
\* 22463 22562: gap of 100 bp  
\* 22563 23720: contig of 1158 bp in length  
\* 23721 23820: gap of 100 bp  
\* 23821 25005: contig of 1185 bp in length  
\* 25006 25105: gap of 100 bp  
\* 25106 26265: contig of 1160 bp in length  
\* 26266 26365: gap of 100 bp  
\* 26366 27528: contig of 1163 bp in length  
\* 27529 27628: gap of 100 bp  
\* 27629 28788: contig of 1160 bp in length  
\* 28789 28888: gap of 100 bp  
\* 28889 30034: contig of 1146 bp in length  
\* 30035 30134: gap of 100 bp  
\* 30135 31237: contig of 1102 bp in length  
\* 31237 31336: gap of 100 bp  
\* 31337 32513: contig of 1176 bp in length  
\* 32513 32612: gap of 100 bp  
\* 32612 33813: contig of 1201 bp in length  
\* 33814 33913: gap of 100 bp  
\* 33914 35107: contig of 1194 bp in length  
\* 35107 35207: gap of 100 bp  
\* 35207 36431: contig of 1224 bp in length  
\* 36431 36531: gap of 100 bp  
\* 36532 37701: contig of 1170 bp in length  
\* 37702 37801: gap of 100 bp  
\* 37802 38967: contig of 1166 bp in length  
\* 38968 39067: gap of 100 bp  
\* 39067 40235: contig of 1168 bp in length  
\* 40236 40335: gap of 100 bp  
\* 40336 41492: contig of 1157 bp in length  
\* 41493 41592: gap of 100 bp  
\* 41593 42691: contig of 1099 bp in length  
\* 42692 42791: gap of 100 bp  
\* 42792 43970: contig of 1179 bp in length  
\* 43971 44070: gap of 100 bp  
\* 44071 45150: contig of 1080 bp in length  
\* 45151 45250: gap of 100 bp  
\* 45251 46347: contig of 1097 bp in length  
\* 46348 46447: gap of 100 bp  
\* 46448 47608: contig of 1161 bp in length  
\* 47609 47708: gap of 100 bp  
\* 47709 48881: contig of 1173 bp in length  
\* 48882 48981: gap of 100 bp  
\* 48982 50240: contig of 1259 bp in length  
\* 50241 50340: gap of 100 bp  
\* 50341 51544: contig of 1204 bp in length  
\* 51545 51644: gap of 100 bp  
\* 51645 52828: contig of 1184 bp in length  
\* 52829 52928: gap of 100 bp  
\* 52929 54157: contig of 1229 bp in length  
\* 54158 54257: gap of 100 bp  
\* 54258 55450: contig of 1199 bp in length  
\* 55451 55550: gap of 100 bp  
\* 55551 56732: contig of 1181 bp in length  
\* 56733 56831: gap of 100 bp  
\* 56832 58019: contig of 1188 bp in length  
\* 58020 58119: gap of 100 bp  
\* 58120 59285: contig of 1166 bp in length  
\* 59286 59385: gap of 100 bp

59386 60525: contig of 1140 bp in length  
\* 60526 60625: gap of 100 bp  
\* 60626 61717: contig of 1092 bp in length  
\* 61718 61817: gap of 100 bp  
\* 61818 62992: contig of 1174 bp in length  
\* 62992 63091: gap of 100 bp  
\* 63092 64267: contig of 1176 bp in length  
\* 64267 64368 65552: contig of 1185 bp in length  
\* 64368 65553 66853: contig of 1201 bp in length  
\* 65553 66954 66953: gap of 100 bp  
\* 66954 68145: contig of 1192 bp in length  
\* 68145 68245: gap of 100 bp  
\* 68245 69424 69523: contig of 1178 bp in length  
\* 69424 69523: gap of 100 bp  
\* 69524 70737: contig of 1214 bp in length  
\* 70738 71997: gap of 100 bp  
\* 70838 71997: contig of 1160 bp in length  
\* 71998 72097: gap of 100 bp  
\* 72098 73254: contig of 1157 bp in length

Query Match 21.3%; Score 27; DB 2; Length 112321;  
Best Local Similarity 100.0%; Pred. No. 4.5e-05;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 CTCGAGGTCGAGGTGATGATGCTCG 45  
Db 102227 CTCGAGGTCGAGGTGATGATGCTCG 102201

RESULT 41  
HSAB323208/c 439 bp DNA linear PRI 18-JUL-2002  
LOCUS HSAB323208  
DEFINITION Homo sapiens genomic sequence surrounding Not1 site, clone  
NB1-073R.  
ACCESSION AJ323208  
VERSION AJ323208.1 GI:15867587  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
AUTHORS Kulesenko, A.S., Giacullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,  
Podowski, R.W., Matushin, Y.G., Gyanchandani, A., Muravenko, O.V.,  
Levitky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I.,  
Kisseliev, L.L., Masserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
TITLE Not1 flanking sequences: a tool for gene discovery and verification  
of the human genome  
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)  
MEDLINE 22131767  
PUBMED 12136098  
REFERENCE 2 (bases 1 to 439)  
AUTHORS Zabarovsky, E.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden

FEATURES  
source 1. 439  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="NB1-073R"

BASE COUNT 63 a 161 c 146 g 68 t 1 others  
ORIGIN

Query Match 19.7%; Score 25; DB 9; Length 439;  
Best Local Similarity 100.0%; Pred. No. 0.00087;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCGAGGTGATG 38

|||||  
Db 217 GCCATCTGCAGGTCGAGGTGATG 193

RESULT 42  
BD096974 706 bp DNA linear PAT 27-AUG-2002  
LOCUS BD096974  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096974  
VERSION BD096974.1 GI:22642562  
KEYWORDS JP 2001526063-A/9.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 706)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 9 18-DEC-2001,  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/9  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 200525451  
PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI  
YI SUN

PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,  
PC A61P39/06,  
PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC  
, C12N5/10, C12N1/68,  
PC G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..291  
FT mat\_peptide 1..291.  
FT Location/Qualifiers  
1. 706  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 189 a 147 c 189 g 181 t  
ORIGIN

Query Match 19.7%; Score 25; DB 6; Length 706;  
Best Local Similarity 100.0%; Pred. No. 0.00087;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCGAGGTGATG 38  
Db 151 GCCATCTGCAGGTCGAGGTGATG 175

RESULT 43  
AC112771 35638 bp DNA linear PRI 28-SEP-2002  
LOCUS AC112771  
DEFINITION Homo sapiens 3 BAC RP11-144C9 (Roswell Park Cancer Institute Human  
BAC Library) complete sequence.  
ACCESSION AC112771  
VERSION AC112771.6 GI:22003982  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 35638)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Albrooks, S.L., Amaratunga, H.C., Are, J.R., Ayala, M., Banks, T.,  
Barbata, U., Benson, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, U., Bowie, S., Briteva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burich, P., Burkett, C., Buttrill, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Bathone, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.U., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C.C., Edgar, D., Edwards, C.C., Elhat, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C.,  
 Hollins, B., Homsi, F., Howard, S., Huber, J., Hultky, S., Hume, J.,  
 Jackson, L.B., Jacobson, B., Jia, Y., Johnson, R., Jolyvet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J.,  
 Kovar, C., Kratovic, J., Kurehi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, U., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louisedag, H., Lozado, R.U., Lu, X., Luckier, A., Luckier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, W., Neal, D.,  
 Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N.,  
 Nicerson, E., Nwokkenwo, S., Oguh, M., Okwomou, G., Oragunye, N.,  
 Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L.,  
 Picens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M.,  
 Rois, A., Rojibokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S.,  
 Scott, G., Shen, H., Shoehat, N., Sisson, I., Sodergren, E.,  
 Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.,  
 Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C.,  
 Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L.,  
 Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S.,  
 Warren, R., Washington, C., Wellington, S., Williams, G.,  
 Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,  
 Wu, Y.P., Zhou, J., Zorrilla, S., Zorrilla, S., Naylor, S.L., Weinstock, G. and  
 Gibbs, R.  
 Title  
 Direct Submission  
 Journal  
 Unpublished  
 Reference  
 2 (bases 1 to 35638)  
 Authors  
 Worley, K.C.  
 Title  
 Direct Submission  
 Journal  
 Submitted (24-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 35638)  
 Authors  
 Worley, K.C.  
 Title  
 Direct Submission  
 Journal  
 Submitted (25-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 35638)  
 Authors  
 Worley, K.C.  
 Title  
 Direct Submission  
 Journal  
 Submitted (30-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 5 (bases 1 to 35638)  
 Authors  
 Worley, K.C.  
 Title  
 Direct Submission  
 Journal  
 Submitted (28-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 30, 2002 this sequence version replaced gi:20976460.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu> or email  
 gc-help@bcm.tmc.edu

ANNOTATION OF FEATURES:  
STSS are identified using ePCR (Genome Res. 7:541-550) searches

```

of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
  Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
  Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.
SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.
QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8086/quality.info/genbank.annotation.ht
ml.
FEATURES
    source
        Location/Qualifiers
            1..35638
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="RP11-144C9"
            1..1895
                /note="overlaps bases 56107..58001 of clone AC092978"
                /function="clone overlap"
            14..341
                /strand_name="SHGC-105132"
                /complement(1283..1580)
                /rpt_family="AluSx"
            1618..1906
                /rpt_family="AluSx"
            1907..1986
                /rpt_family="AluSx"
            /rpt_family="GGGGA)n"
                /complement(2109..2203)
                /rpt_family="L1M4"
            /complement(2204..2463)
                /rpt_family="AluSx"
            /complement(2464..2700)
                /rpt_family="L1M4"
            /complement(2701..2819)
                /rpt_family="AluSc"
            /complement(2843..3539)
                /rpt_family="L1M4"
            /complement(3541..3642)
                /rpt_family="AluSg/x"
            3643..3758
                /rpt_family="FLM_A"
                /complement(3777..4993)
                /rpt_family="L1MEC"
            /complement(4994..5287)
                /rpt_family="AluSg"
            /complement(5288..5404)
                /rpt_family="L1MEC"
            /complement(5405..5716)
                /rpt_family="AluSx"
            /complement(5717..6153)
                /rpt_family="L1MEC"
            /complement(6154..6379)
                /rpt_family="AluSg/x"
            /complement(6380..6394)
                /rpt_family="L1MEC"
            /complement(6427..6570)
                /rpt_family="MER91A"

```

```

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
  http://www.hgsc.bcm.tmc.edu:8008/quality.info/genbank.annotation.ht
  ml.

CURS
source
  location/Qualifiers
  1..35638
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="3"
  /clone="RP11-144C9"
  1..1895
  /note="overlaps bases 56107..58001 of clone AC092978"
  /function="clone overlap"
  14..341
  /standard_name="SHGC-105132"
  complement(1283..1580)
  /rpt_family="AluXs"
  1618..1906
  /rpt_family="AluXs"
  1907..1986
  /rpt_family="GGGGA)n"
  complement(2109..2203)
  /rpt_family="L1M4"
  complement(2204..2463)
  /rpt_family="AluXs"
  complement(2464..2700)
  /rpt_family="L1M4"
  complement(2701..2819)
  /rpt_family="AluSc"
  complement(2843..3359)
  /rpt_family="L1M4"
  complement(3541..3642)
  /rpt_family="AluSg/x"
  3643..3758
  /rpt_family="FLAM A"
  complement(3777..4993)
  /rpt_family="L1MEC"
  complement(4994..5287)
  /rpt_family="AluSg"
  complement(5288..5404)
  /rpt_family="L1MEC"
  complement(5405..5716)
  /rpt_family="AluXs"
  complement(5717..6153)
  /rpt_family="L1MEC"
  complement(6154..6379)
  /rpt_family="AluSg/x"
  complement(6380..6394)
  /rpt_family="L1MEC"
  complement(6427..6570)
  /rpt_family="MER91A"

```

```

repeat_region      7418..7561
                    /rpt_family="AluSg/x"
repeat_region      8094..8139
                    /rpt_family="(TG)n"
repeat_region      complement(10186..10802)
                    /rpt_family="L2"
repeat_region      11949..12489
                    /rpt_family="HAL1"
repeat_region      12490..12772
                    /rpt_family="AluSx"
repeat_region      12773..12804
                    /rpt_family="(TAAAA)n"
repeat_region      12805..12820
                    /rpt_family="AluSx"
repeat_region      12821..12896
                    /rpt_family="HAL1"
repeat_region      12989..13063
                    /rpt_family="AluSg/x"
repeat_region      13069..13452
                    /rpt_family="HAL1"
repeat_region      13453..13585
                    /rpt_family="FLAM_C"
repeat_region      13586..13609
                    /rpt_family="HAL1"
repeat_region      13619..13654
                    /rpt_family="G-rich"
repeat_region      13653..13809
                    /rpt_family="G-rich"
repeat_region      complement(13819..14048)
                    /rpt_family="AluJo"
repeat_region      14077..14170
                    /rpt_family="MER113"
repeat_region      complement(14655..14830)
                    /rpt_family="MUT1U"
repeat_region      complement(14852..15139)
                    /rpt_family="AluSp"

```

```

Query Match      19.7% Score 25; DB 9; Length 35638;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      14 GCCATCTGCAGGTCCTCAGGTGATGG 38
Db      25655 GCCATCTGCAGGTCCTCAGGTGATGG 25679

```

```

RESULT 44
AC106176/c      238330 bp      DNA      linear      HTG 13-MAY-2003
LOCUS      Rattus norvegicus clone CH230-20P16, WORKING DRAFT SEQUENCE, 3
DEFINITION      unnumbered pieces.
ACCESSION      AC106176
VERSION      AC106176.5 GI:30578781
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus

```

```

REFERENCE
AUTHORS      1 (bases 1 to 238330)
              Muzny,D,Marie,, Metzker,M,Lee,, Abramson,S,, Adams,C,, Alder,J,,
              Allen,C,, Allen,H,, Alspbrooks,S,, Amin,A,, Anguiano,D,,
              Anyalelechi,V,, Aoyagi,A,, Ayodeji,M,, Baca,E,, Baden,H,,
              Baldwin,D,, Bandaranaike,D,, Barber,M,, Barnstead,M,, Benahmed,F,,
              Bisselt,K,, Blair,J,, Blankenburg,K,, Blyth,P,, Brown,M,,
              Bryant,N,, Buhay,C,, Burch,P,, Burrell,K,, Calderon,E,,
              Cardenas,V,, Carter,K,, Cavazos,I,, Ceasar,H,, Center,A,,
              Chacko,J,, Chavez,D,, Chen,G,, Chen,R,, Chen,Y,, Chen,Z,, Chu,J,,
              Cleveland,C,, Cockrell,R,, Cox,C,, Coyle,M,, Cree,A,, P'Souza,L,,
              Davila,M.L., Davis,C,, Davy-Carrillo,L,, De Anda,C,, Dedertich,D,,
              Delgado,O,, Denison,S,, Deramo,C,, Ding,Y,, Dinh,C,, Divya,K,,
              Draper,H,, Dugan-Rocha,S,, Dunn,A,, Durbin,K,, Duval,B,, Eaves,K,,
              Egan,A,, Escotto,M,, Eugene,C,, Evans,C.A., Falls,T,, Fan,G,,

```

```

TITLE      JOURNAL
REFERENCE   JOURNAL
AUTHORS     Worley,K.C.
TITLE      Direct Submission
JOURNAL     Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 238330)
            Rat Genome Sequencing Consortium.
            Direct Submission
            Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On May 13, 2003 this sequence version replaced gi:24819215.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.

```

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUOH
Center clone name: CH230-20P16
----- Summary Statistics
Assembly program: Atlas 3.0;

```

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,B., Geer,K., Gill,R., Grady,A., Guerra,M., Guevara,M., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huylk,S., Hune,J., Idlibird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Loulseged,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundaas,M., Murphy,M., Natr,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwackemehe,O., Okwomu,G., Olarunsaagoun,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindester,A., Popovic,D., Primus,E., Pu,J., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wiczzyk,R., Wooden,K., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

```

Consensus quality: 229461 bases at least Q40
Consensus quality: 231714 bases at least Q30
Consensus quality: 233148 bases at least Q20
Estimated insert size: 239075; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 236056: contig of 236056 bp in length
* 236057 236156: gap of unknown length
* 236157 237213: contig of 1057 bp in length
* 237214 237313: gap of unknown length
* 237314 238330: contig of 1017 bp in length.
*
* Location/Qualifiers
*     1..238330
*         /organism="Rattus norvegicus"
*         /mol_type="genomic DNA"
*         /db_xref="taxon:10116"
*         /clone="CH230-20P16"
*         1..1582
*         /note="vgs contig"
misc_feature 64585 a 53668 c 53820 g 62326 t 3931 others
BASE COUNT
ORIGIN

Query Match 19.7%; Score 25; DB 2; Length 238330;
Best Local Similarity 100.0%; Pred. No. 0.00078;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGAGGGTCAGGTGATGG 38
DB 164432 GCCATCTGAGGGTCAGGTGATGG 164408

RESULT 45
AC098496/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-23C21, WORKING DRAFT SEQUENCE, 4
unmerged pieces.
ACCESSION AC098496
VERSION AC098496.8 GI:30521845
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 245476)
Muzny, D. Marie, Metzker, M. Lee, Abramzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsdorfs, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Dayala, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Dejgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Driper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorjais, B., Geer, K., Gilly, R., Grady, M., Guerra, W., Guevara, W.,
Guanarane, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenzenhwa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,
Mansueti, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Manhley, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Mnjia, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Markovic, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokolemeh, O., Okunolu, G., Olarunsaogun, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, J., Plamoch, C.,
Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quirroz, J., Rachin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Raves, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savery, G., Scherer, S., Scott, C., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 245476)
Worley, K. C.
Direct Submission
Submitted (24-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 245476)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24819573.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHEY
Center clone name: CH230-23C21
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 228381 bases at least Q40
Consensus quality: 230024 bases at least Q30
Consensus quality: 231602 bases at least Q20
Estimated insert size: 236545; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

```



```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5516: contig of 5516 bp in length
* 5517 5616: gap of unknown length
* 5617 242823: contig of 237207 bp in length
* 242824 242923: gap of unknown length
* 242924 243958: contig of 1035 bp in length
* 243959 244058: gap of unknown length
* 244059 245476: contig of 1418 bp in length.
* Location/Qualifiers
*     1. 245476
*        /organism="Rattus norvegicus"
*        /mol_type="genomic DNA"
*        /db_xref="taxon:10116"
*        /clone="CH230-23C21"
*        1. 1560
*        /note="wgs_end_extension"
*        clone_end:17
*        complement(3889..4558)
*        /note="clone_boundary"
*        clone_end:17
*        site:ECORI
*        end sequence: BH361779"
* 5617..6729
*        /note="wgs_contig"
*        69231..103385
*        /note="clone_boundary"
*        clone_end:Sp6
*        site:ECORI
*        end sequence: BH361780"
*
BASE COUNT 67659 a 50656 c 50781 g 63728 t 12652 others
ORIGIN
Query Match 19.7%; Score 25; DB 2; Length 245476;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GCCATCTGCAGGTCGAGTGATGG 38
Db 69519 GCCATCTGCAGGTCGAGTGATGG 69495
RESULT 46
AC118520/c 189541 bp DNA linear HTG 20-NOV-2002
LOCUS Rattus norvegicus clone CH230-304J21, *** SEQUENCING IN PROGRESS
DEFINITION
***, 2 unordered pieces.
AC118520
AC118520.4 GI:25138134
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 189541)
Muzny,D,Marle,M, Metzker,M, Lee,A, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,

```

```

Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Gebregorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunaratne,P, Haaland,W, Hamli,C, Hamilton,C, Hamilton,K,
Harvey,Y, Haylak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howells,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Liu,J,
Kowis,C, Kratt,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorensuhewa,L, Loulseged,H, Lozada,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwackemele,O, Okwuonu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
Plopper,F, Polindexter,A, Popovic,D, Prims,E, Pu,L,
Puato,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reich,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Rutz,S,J,
Sanders,M, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
Shetty,J, Shvartbeyn,A, Sisson,I, Sitter,C,D, Smajs,D,
Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usman,K,
Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J,
Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
Williams,G, Willson,R, Wleczkyk,R, Wooden,H, Worley,K,
Wright,D, Wright,R, Wu,J, Yakub,S, Yen,T, Yoon,L, Yoon,V,
Yu,P, Zhang,U, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von
Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
Weinstock,G, and Gibbs,R,A.
Direct Submission
Unpublished
2 (bases 1 to 189541)
Worley,K,C.
Direct Submission
Submitted (18-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 189541)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:33914484.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```



NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.bjsg.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.bjsg.bcm.tmc.edu/docs/genbank/draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 52 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of 'N', but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 163: contig of 1163 bp in length  
 1164 1263: gap of unknown length  
 1264 2611: contig of 1348 bp in length  
 2612 2712: gap of unknown length  
 2712 3837: contig of 1125 bp in length  
 3837 3936: gap of unknown length  
 3936 5317: contig of 1381 bp in length  
 5317 5417: gap of unknown length  
 5418 6568: contig of 1151 bp in length  
 6568 7872: gap of unknown length  
 7872 7972: contig of 1204 bp in length  
 7972 9049: gap of unknown length  
 9049 9149: contig of 1077 bp in length  
 9149 10503: contig of 1354 bp in length  
 10504 10603: gap of unknown length  
 10604 11740: contig of 1137 bp in length  
 11741 11840: gap of unknown length  
 11841 12996: contig of 1156 bp in length  
 12997 13097: gap of unknown length  
 13097 14621: contig of 1525 bp in length  
 14622 14721: gap of unknown length  
 14722 15860: contig of 1139 bp in length  
 15861 15960: gap of unknown length  
 15961 17457: contig of 1497 bp in length  
 17458 17557: gap of unknown length  
 17558 19168: contig of 1611 bp in length  
 19169 19268: gap of unknown length  
 19269 20749: contig of 1481 bp in length  
 20750 20849: gap of unknown length  
 20850 22819: contig of 1970 bp in length  
 22820 22919: gap of unknown length  
 22920 25192: contig of 2273 bp in length  
 25193 25292: gap of unknown length  
 25293 27234: contig of 1942 bp in length  
 27235 27335: gap of unknown length  
 27335 28926: contig of 1592 bp in length  
 28927 29026: gap of unknown length  
 29027 30173: contig of 1147 bp in length  
 30174 30273: gap of unknown length  
 30274 32443: contig of 2170 bp in length  
 32444 32543: gap of unknown length  
 32544 34092: contig of 1549 bp in length  
 34093 34192: gap of unknown length  
 34193 37288: contig of 3096 bp in length  
 37289 37388: gap of unknown length  
 37389 40619: contig of 3231 bp in length  
 40620 40719: gap of unknown length  
 40720 43147: contig of 2428 bp in length  
 43148 43247: gap of unknown length  
 43248 45341: contig of 2094 bp in length  
 45342 45441: gap of unknown length  
 45442 49153: contig of 3712 bp in length  
 49154 49253: gap of unknown length  
 49254 52591: contig of 3338 bp in length  
 52592 52691: gap of unknown length  
 52692 55192: contig of 2401 bp in length  
 55193 55674: gap of unknown length  
 55675 59774: contig of 4482 bp in length  
 59775 62517: gap of unknown length  
 62518 62617: gap of unknown length

62618 65416: contig of 2799 bp in length  
 65417 65516: gap of unknown length  
 65517 69364: contig of 3848 bp in length  
 69365 69464: gap of unknown length  
 69465 73440: contig of 3976 bp in length  
 73441 73540: gap of unknown length  
 73541 77575: contig of 4035 bp in length  
 77576 77675: gap of unknown length  
 77676 82110: contig of 4435 bp in length  
 82111 82210: gap of unknown length  
 82211 87479: contig of 5269 bp in length  
 87480 87579: gap of unknown length  
 87580 93557: contig of 5978 bp in length  
 93558 93657: gap of unknown length  
 93658 98254: contig of 4597 bp in length  
 98255 98354: gap of unknown length  
 98355 104083: contig of 5729 bp in length  
 104084 104183: gap of unknown length  
 111520 111520: contig of 7337 bp in length  
 111521 111620: gap of unknown length  
 111621 116795: contig of 5075 bp in length  
 116796 116796: gap of unknown length  
 116797 123174: contig of 6379 bp in length  
 123175 123274: gap of unknown length  
 123275 131678: contig of 8404 bp in length  
 131679 131778: gap of unknown length  
 131779 138679: contig of 6901 bp in length  
 138680 138779: gap of unknown length  
 138780 146691: contig of 7912 bp in length  
 146692 146791: gap of unknown length  
 146792 153400: contig of 6609 bp in length  
 153401 153500: gap of unknown length  
 153501 161119: contig of 7619 bp in length  
 161120 161219: gap of unknown length  
 161220 161220: contig of 10537 bp in length

Query Match 18.9% Score 24; DB 2; Length 208141;  
 Best Local Similarity 100.0%; Pred. No. 0.0032;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 76 GGACTGTGTTGCTGGGAGAGA 99  
 Db 12567 GGACTGTGTTGCTGGGAGAGA 12544

RESULT 48  
 AC094582/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-4M15, WORKING DRAFT SEQUENCE.  
 AC094582.7 GI:30466356  
 ACCESSION  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 1 (bases 1 to 243487)  
 Muzny D, Marie M, Metker M, Lee A, Abramson S, Adams C, Alder J, Allen C, Allen H, Alsbrooks S, Amn A, Angiano D, Anyalelechi V, Aoyagi A, Ayodeji M, Baca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F, Biswal K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon E, Cardenas V, Carter K, Cavazos I, Ceaar H, Center A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L, Davila M, Davis C, Davy-Carroll L, De Anda C, Dederich D, Delgado O, Denison S, Deramo C, Ding R, Dinn H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Evans K, Egan A, Escotto M, Eugene C, Evans C, A, Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Fraser C, M., Gabisi A, Ganta R, Garcia A, Garner T, Garza M,

Guebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M.  
Gunnatre, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, T.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hult, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Louieged, H., Lozada, R. J., Lu, X., Ma, D.,  
Maheshwari, M., Mahindratne, M., Mahmoud, M., Mallory, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nait, L.,  
Nankrevits, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwakoelameh, O., Okwionu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoth, C.,  
Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, B., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rutz, S. T.,  
Sanders, W., Savary, G., Scheerer, S., Scott, G., Shatsman, S., Shen, H.,  
Sherry, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodegren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabori, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlaczky, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Wiederstaedt, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G., and Gibbs, R. A.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 243487)
AUTHORS	Morley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 243487) Rat Genome Sequencing Consortium. Direct Submission.
AUTHORS	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	On May 9, 2003 this sequence version replaced gi:24318567.
COMMENT	

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: GAYK  
Center clone name: CH230-4M15  
----- Summary Statistics  
Assembly program: APLAS  
Consensus quality: 216410 bases at least Q40  
Consensus quality: 218103 bases at least Q30

```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 243487: contig of 243487 bp in length.

```

```
FEATURES
    source
        Location/Qualifiers
            1..243487
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10116"
                /clone="CH230-4M15"
            1..2254
                /note="wgs end_extension
                    clone_end:T7"
                complement(3914..4636)
                /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH309319"
            235713..236568
                /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH309319"
            242181..243487
                /note="wgs end_extension
                    clone_end:Sp6"
```

Query Match	18.9%	Score 24;	DB 2;	Length 243487;
Best Local Similarity	100.0%	Pred. No. 0.0032;		
Matches 24;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy 76 GGACTGCTTGTGGTGGGAGA 99  
 |||||  
 Db 1773 GGACTGCTTGTGGTGGGAGA 17710

LOCUS	AC104983	59155 bp	DNA	linear	HTG 22-DEC-2001
DEFINITION	Homo sapiens chromosome 18 clone RP11-352C3 map 18, LOW-PASS SEQUENCE SAMPLING.				

ACCESSION	AC104983
VERSION	AC104983.1
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 59155)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 18, clone RP11-352C3  
Unpublished  
2 (bases 1 to 59155)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., B...

Eukaryota; euteleostomi; Choroata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 59155)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 18, clone RP11-352C3  
Unpublished  
2 (bases 1 to 59155)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barta, N., Bastien, V., Boguski, V., Bouckgeater, B., Brown, A., Camarata, J., Campione, A., Chang, Y., Chazaro, B., Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearfano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde,S., Gord,S., Coyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamazaras,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,  
Maclean,C., MacDonald,P., Major,J., Margis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPeckers,R., Meldrum,J.,  
Menas,L., Milova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Romanov,I., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpack,R.,  
Seaman,S., Severy,P., Spencer,P., Stange-Thomann,N., Stojanovic,N.,  
Struss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggilo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zaitoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L22883  
Center clone name: 352\_C\_3  
-----

\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
725 724: contig of 724 bp in length  
825 824: gap of 100 bp  
1583 1582: contig of 758 bp in length  
1683 1682: gap of 100 bp  
2409: contig of 727 bp in length  
2509: gap of 100 bp  
2510 3235: contig of 726 bp in length  
3236 3335: gap of 100 bp  
3336 4090: contig of 755 bp in length  
4091 4190: gap of 100 bp  
4191 4955: contig of 765 bp in length  
4956 5055: gap of 100 bp  
5056 5780: contig of 725 bp in length  
5781 5880: gap of 100 bp  
5881 6643: contig of 763 bp in length  
6644 6743: gap of 100 bp  
6744 7496: contig of 753 bp in length  
7497 7596: gap of 100 bp  
7597 8348: contig of 752 bp in length  
8349 8448: gap of 100 bp  
8449 9217: contig of 769 bp in length  
9218 9317: gap of 100 bp  
9318 10093: contig of 776 bp in length  
10094 10193: gap of 100 bp  
10194 10924: contig of 731 bp in length  
10925 11024: gap of 100 bp  
11025 11793: contig of 769 bp in length  
11794 11893: gap of 100 bp  
11894 12661: contig of 768 bp in length  
12662 12761: gap of 100 bp  
12762 13525: contig of 764 bp in length

13526 13625: gap of 100 bp  
13626 14347: contig of 722 bp in length  
14348 14447: gap of 100 bp  
14448 15172: contig of 725 bp in length  
15173 15272: gap of 100 bp  
15273 16007: contig of 735 bp in length  
16008 16107: gap of 100 bp  
16108 16855: contig of 748 bp in length  
16856 16955: gap of 100 bp  
16956 17711: contig of 736 bp in length  
17712 17811: gap of 100 bp  
17812 18568: contig of 757 bp in length  
18569 18668: gap of 100 bp  
18669 19421: contig of 753 bp in length  
19422 19521: gap of 100 bp  
19522 20229: contig of 708 bp in length  
20230 20329: gap of 100 bp  
20330 21061: contig of 732 bp in length  
21062 21161: gap of 100 bp  
21162 21909: contig of 748 bp in length  
21910 22009: gap of 100 bp  
22010 22773: contig of 764 bp in length  
22774 22873: gap of 100 bp  
22874 23623: contig of 750 bp in length  
23624 23723: gap of 100 bp  
23724 24489: contig of 766 bp in length  
24490 24589: gap of 100 bp  
24590 25354: contig of 765 bp in length  
25355 25454: gap of 100 bp  
25455 26174: contig of 720 bp in length  
26175 26274: gap of 100 bp  
26275 27043: contig of 769 bp in length  
27044 27143: gap of 100 bp  
27144 27900: contig of 757 bp in length  
27901 28000: gap of 100 bp  
28001 28753: contig of 753 bp in length  
28754 28853: gap of 100 bp  
28854 29572: contig of 719 bp in length  
29573 29672: gap of 100 bp  
29673 30419: contig of 747 bp in length  
30420 30519: gap of 100 bp  
30520 31262: contig of 743 bp in length  
31263 31362: gap of 100 bp  
31363 32052: contig of 690 bp in length  
32053 32152: gap of 100 bp  
32153 32909: contig of 757 bp in length  
32910 33009: gap of 100 bp  
33010 33736: contig of 727 bp in length  
33737 33836: gap of 100 bp  
33837 34598: contig of 762 bp in length  
34599 34698: gap of 100 bp  
34699 35456: contig of 758 bp in length  
35457 35556: gap of 100 bp  
35557 36274: contig of 718 bp in length  
36275 36374: gap of 100 bp  
36375 37111: contig of 737 bp in length  
37112 37211: gap of 100 bp  
37212 37954: contig of 743 bp in length  
37955 38054: gap of 100 bp  
38055 38783: contig of 729 bp in length  
38784 38883: gap of 100 bp  
38884 39686: contig of 803 bp in length  
39687 39786: gap of 100 bp  
39787 40535: contig of 749 bp in length  
40536 40635: gap of 100 bp  
40636 41387: contig of 753 bp in length  
41388 41487: gap of 100 bp  
41488 42254: contig of 767 bp in length  
42255 42354: gap of 100 bp  
42355 43046: contig of 692 bp in length  
43047 43146: gap of 100 bp  
43147 43885: contig of 739 bp in length  
43886 43985: gap of 100 bp

TITLE  
JOURNAL  
COMMENT

```

* 43986 44741: contig of 756 bp in length
* 44742 44841: gap of 100 bp
* 44842 44841: contig of 763 bp in length
* 44842 45604: contig of 763 bp in length
* 45605 45704: gap of 100 bp
* 45704 46429: contig of 725 bp in length
* 46429 46430: gap of 100 bp
* 46430 46530: contig of 715 bp in length
* 46530 47245: gap of 100 bp
* 47245 47345: gap of 100 bp
* 47345 48088: contig of 744 bp in length
* 48088 48189: gap of 100 bp
* 48189 48948: contig of 760 bp in length
* 48948 49049: gap of 100 bp
* 49049 49803: contig of 755 bp in length
* 49803 49903: gap of 100 bp
* 49903 50649: contig of 745 bp in length
* 50649 50748: gap of 100 bp
* 50748 51506: contig of 758 bp in length
* 51506 51507: gap of 100 bp
* 51507 52345: contig of 739 bp in length
* 52345 52445: gap of 100 bp
* 52445 53185: contig of 740 bp in length
* 53185 53285: gap of 100 bp
* 53285 54059: contig of 774 bp in length
* 54059 54159: gap of 100 bp
* 54159 54901: contig of 742 bp in length
* 54901 55001: gap of 100 bp
* 55001 55740: contig of 739 bp in length
* 55740 55840: gap of 100 bp
* 55840 56577: contig of 737 bp in length
* 56577 56678: gap of 100 bp
* 56678 57433: contig of 756 bp in length
* 57433 57534: gap of 100 bp
* 57534 58290: contig of 757 bp in length

```

Query Match 17.3%; Score 22; DB 2; Length 59155;  
 Best Local Similarity 100.0%; Pred. No. 0.058;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGTCT 92  
 |||||

Db 7798 CAAGAGACTGTGTGTGTCT 7819

RESULT 50  
 AC108040/c  
 LOCUS Homo sapiens chromosome 4 clone RP11-210010, WORKING DRAFT  
 DEFINITION AC108040  
 AC108040  
 AC108040.2 GI:18425316  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILLTOP.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 163521)  
 Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 163521)  
 AUTHORS Waterston, R.H.  
 JOURNAL Direct Submission  
 TITLE Submitted (24-JAN-2002) Genome Sequencing Center, Washington  
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On Jan 30, 2002 this sequence version replaced gi:18308830.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@watson.wustl.edu  
 ----- Project Information -----

Center project name: H.NH0210010  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-Primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 162304 bases at least Q40  
 Consensus quality: 162444 bases at least Q30  
 Consensus quality: 162529 bases at least Q20  
 Insert size: 185000; agarose-fp  
 Insert size: 163221; sum-of-contigs  
 Quality coverage: 8.88 in Q20 bases; agarose-fp  
 Quality coverage: 8.85 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1564: contig of 1564 bp in length  
 \* 1565 1664: gap of unknown length  
 \* 1665 88984: contig of 87320 bp in length  
 \* 88985 89084: gap of unknown length  
 \* 89085 112702: contig of 23618 bp in length  
 \* 112703 112803: gap of unknown length  
 \* 112803 163521: contig of 50719 bp in length.  
 FEATURES  
 source  
 1. 163521  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /clone="RP11-210010"  
 1. 1564  
 /note="assembly\_name:Contig25"  
 clone\_end:SP6  
 vector\_side:left  
 1665\_88984  
 /note="assembly\_name:Contig28"  
 clone\_end:T7  
 vector\_side:left  
 89085\_112702  
 /note="assembly\_name:Contig26"  
 112803\_163521  
 /note="assembly\_name:Contig27"  
 BASE COUNT 51049 a 32426 c 33346 g 46399 t 301 others  
 ORIGIN

Query Match 17.3%; Score 22; DB 2; Length 163521;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGTCT 92  
 |||||

Db 80211 CAAGAGACTGTGTGTGTCT 80190

Search completed: November 7, 2003, 13:29:05  
 Job time : 847.956 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 02:39:56 ; Search time 129.53 Seconds  
(without alignments)  
2646.715 Million cell updates/sec

Title: US-09-509-779-1\_COPY\_154\_280

Perfect score: 127  
Sequence: 1 GTGGCATACCTGTGCTCATCT.....ATTCTTCACACACGTCTGC 127

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq\_190N03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	100.0	1140	20	AAx87313
2	32	25.2	224	20	AAx41085
3	32	25.2	342	21	AAx96883
4	32	25.2	342	22	AAH43570
5	32	25.2	342	24	AAH39674
6	32	25.2	754	20	AAx87317
7	32	25.2	754	20	AAx87321
8	32	25.2	754	20	AAx87322

9	32	25.2	754	20	AAx87323	Human sensitive to
10	32	25.2	754	20	AAx87324	Human sensitive to
11	32	25.2	754	20	AAx87325	Human sensitive to
12	32	25.2	754	20	AAx87326	Human sensitive to
13	32	25.2	754	20	AAx87327	Human sensitive to
14	32	25.2	754	20	AAx87328	Human sensitive to
15	32	25.2	754	20	AAx87329	Human sensitive to
16	32	25.2	754	20	AAx87330	Human sensitive to
17	32	25.2	754	20	AAx87331	Human sensitive to
18	32	25.2	754	20	AAx87332	Human sensitive to
19	32	25.2	836	22	AAx87333	Human sensitive to
20	32	25.2	836	25	AAx87334	Human sensitive to
21	32	25.2	836	25	AAx87335	Human sensitive to
22	32	25.2	962	21	AAx87336	Human sensitive to
23	30	23.6	754	20	AAx87337	Human sensitive to
24	30	23.6	754	20	AAx87338	Human sensitive to
25	29	22.8	747	20	AAx87339	Human sensitive to
26	29	22.8	754	20	AAx87340	Human sensitive to
27	29	22.8	754	20	AAx87341	Human sensitive to
28	29	22.8	1152	22	AAx87342	Human sensitive to
29	25	19.7	60	24	AAx87343	Human sensitive to
30	25	19.7	439	24	AAx87344	Human sensitive to
31	25	19.7	441	21	AAx87345	Human sensitive to
32	25	19.7	706	20	AAx87346	Human sensitive to
33	21	16.5	264	22	AAx87347	Human sensitive to
34	21	16.5	596	22	AAx87348	Human sensitive to
35	21	16.5	630	22	AAx87349	Human sensitive to
36	19	15.0	225	24	AAx87350	Human sensitive to
37	19	15.0	498	22	AAx87351	Human sensitive to
38	19	15.0	498	22	AAx87352	Human sensitive to
39	18	14.2	18	20	AAx87353	Human sensitive to
40	17	13.4	239	23	AAx87354	Human sensitive to
41	17	13.4	348	25	AAx87355	Human sensitive to
42	17	13.4	355	22	AAx87356	Human sensitive to
43	17	13.4	355	22	AAx87357	Human sensitive to
44	17	13.4	355	22	AAx87358	Human sensitive to
45	17	13.4	355	22	AAx87359	Human sensitive to

#### ALIGNMENTS

RESULT 1	AAx87313	standard; cDNA, 1140 BP.
AC	AAx87313;	
XX		
DT	27-SEP-1999	(first entry)
XX		
DE	Mouse sensitive to apoptosis (SAG) gene.	
XX		
KW	SAG gene; sensitive to apoptosis; mouse; cancer; tumour;	
KW	neurodegenerative disease; muscular dystrophy; wound healing;	
XX	vulnerary; therapy; ds.	
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	17..358
FT		/*tag= a
XX		
FN	W09932514-A2.	
XX		
PD	01-JUL-1999.	
XX		
PF	15-DEC-1998;	98WO-US26705.
XX		
PR	11-SEP-1998;	98US-0099849
XX	19-DEC-1997;	97US-0068179
XX		
PA	(WARN ) WARNER LAMBERT CO.	
XX		

```

PI Sun Y;
XX
XX WPI: 1999-430152/36.
DR P-PSDB; AAY06491.
XX
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for
PT promoting cell growth and protecting cells against apoptosis
XX
PS Claim 1; Page 48-49; 84pp; English.
XX
XX This is the nucleotide sequence of mouse sensitive to apoptosis
CC (SAG) cDNA, which codes for a novel redox-sensitive, haem-binding
CC protein (see AAY06491) with a zinc RING finger domain that promotes
CC cell growth, protects cells from apoptosis, scavenges oxygen
CC radicals and can be used for the reversal of a tumour phenotype.
CC The cDNA was isolated using differential display to identify genes
CC associated with 1,10-phenanthroline (OP)-induced apoptosis in
CC murine tumour lines, and use of an isolated OP-inducible clone to
CC screen a mouse lung cDNA library. The mouse SAG cDNA was used to
CC identify human SAG (see AAX87314). SAG is highly conserved among
CC species. Disruption in yeast was shown to be lethal. SAG deletion
CC mutants (see AAX87315-16) have been identified in human cancer lines,
CC suggesting a role in carcinogenesis. SAG genes, and mutant SAG
CC genes, can be used to protect cells from apoptosis induced by redox
CC reagents. Antisense SAG genes can be used to inhibit the growth of
CC tumour cells. The SAG genes can also be used for the recombinant
CC production of the SAG proteins. The SAG proteins can be used to
CC scavenge oxygen radicals in organisms and to promote wound healing.
CC Additionally, the SAG genes or their complements can be used to
CC promote or inhibit the growth of plant cells (all claimed). The SAG
CC protein is also an ideal molecular target in the development of
CC drugs against neurodegenerative disorders, cancers and muscle
CC dystrophy.
XX
XX Sequence 1140 BP; 302 A; 224 C; 287 G; 327 T; 0 other;
SQ
Query Match 100.0%; Score 127; DB 20; Length 1140;
Best Local Similarity 100.0%; Pred. No. 2.5e-57;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTGCGATACCTGTGCGATCTGCAGGGTCCAGGTGATGCGCTTCGATGCAAGC 60
DB 154 GTGCGATACCTGTGCGATCTGCAGGGTCCAGGTGATGCGCTTCGATGCAAGC 213
OY 61 TGAANAACAACAAGGAGCTGTGTGTGTCTGCGGAGAAGTGAACCATTCCTCCACAA 120
DB 214 TGAANAACAACAAGGAGCTGTGTGTGTGTCTGCGGAGAAGTGAACCATTCCTCCACAA 273
OY 121 CTGCTGC 127
DB 274 CTGCTGC 280

```

```

XX
XX 11-FEB-1999.
PD
XX
XX 31-JUL-1998; 98WD-IB01238.
PF
XX
XX 01-AUG-1997; 97US-0905134.
PR
XX
XX (GENSET) GENSET.
PA
XX
XX Duclert A, Dumas Milne Edwards J, Lacroix B;
PI
XX WPI: 1999-153784/13.
DR
XX P-PSDB; AAY12252.
XX
XX New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Claim 1; Page 441; 622pp; English.
XX
XX AAY40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAY01602 and
XX AAY11994 to AAY12260, respectively. The proteins given represent the
XX signal peptide and an N-terminal fragment of a secreted protein. The
XX nucleic acid sequences can be used for producing secreted human gene
XX products. They can also be used to develop products for diagnosis and
XX therapy. The proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used
XX for directing extracellular secretion of a polypeptide or the insertion
XX of a polypeptide into a membrane, or importing a polypeptide into
XX a cell.
SQ
Sequence 224 BP; 41 A; 60 C; 83 G; 38 T; 2 other;
Query Match 25.2%; Score 32; DB 20; Length 224;
Best Local Similarity 100.0%; Pred. No. 4.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 14 GCCATCTGCAGGGTCCAGGTGATGATGCTG 45
DB 174 GCCATCTGCAGGGTCCAGGTGATGATGCTG 205

```

```

RESULT 2
AAX41085
ID AAX41085 standard; cDNA; 224 BP.
XX
XX AAX41085;
AC
XX
XX 18-JUN-1999 (first entry)
DT
XX
XX Human secreted protein 5' EST SEQ ID NO: 297.
DE
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition; ds.
KM
XX Homo sapiens.
OS
XX
XX MO9906554-A2.
PN

```

```

RESULT 3
AAA96883
ID AAA96883 standard; DNA; 342 BP.
XX
XX AAA96883;
AC
XX
XX 19-FEB-2001 (first entry)
DT
XX
XX Nucleotide sequence of human ring finger protein ROC2.
DE
XX
XX ROC1; ROC2; cullin; ring finger protein; APC11; APC complex; SCF pathway;
XX cullin dependent ubiquitin ligase; CDK inhibitor Sic1 degradation;
XX tumour; ss.
KM
XX
XX Homo sapiens.
OS
XX
XX Key CDS
FH
XX 1..342
FT
XX
XX /tag= a
FT
XX /product= "ROC2"
FT
XX /transl_except= (pos: 28..30, aa: Pro)
FT
XX /transl_except= (pos: 34..36, aa: Val)
FT
XX /transl_except= (pos: 40..42, aa: Ser)

```



```

FT      /cransl_except= (pos: 58..60, aa: Ala)
FT      /cransl_except= (pos: 67..69, aa: Lys)
FT      /cransl_except= (pos: 109..111, aa: Ala)
XX
XX      MO200058472-A2.
XX
XX      05-OCT-2000.
XX
XX      31-MAR-2000; 2000MO-US08592.
XX
XX      31-MAR-1999; 99US-0127261.
XX      22-NOV-1999; 99US-0166927.
XX
XX      (UYNC-) UNIV NORTH CAROLINA.
XX
XX      Xiong Y, Ohta T;
XX
XX      WPI; 2000-647235/62.
XX      P-PSDB; AAB19161.
XX
XX      Novel nucleic acid encoding cullin regulating ring finger proteins,
XX      termed as ROC proteins similar to anaphase-promoting complex 11, for
XX      therapeutic and diagnostic use
XX
XX      Claim 18; Fig 2B; 83pp; English.
XX
XX      The present sequence encodes a human ROC2 ring finger protein. The
XX      specification also describes human ROC1, ROC1 and ROC2 are similar
XX      to APC11, a subunit of the APC complex. The proteins stimulate cullin
XX      dependent ubiquitin ligase activity. ROC1 functions in vivo as an
XX      essential regulator of CDK inhibitor Sic1 degradation by the SCF
XX      (undefined) pathway. ROC proteins are useful for screening bioactive
XX      agents that interfere with the binding of ROC proteins with cullin
XX      proteins. Pharmaceutical formulations comprising ROC proteins are
XX      useful for diagnostic and therapeutic purposes, preferably for
XX      diagnosing and treating tumours.
XX
XX      Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other;
XX
XX      Query Match      25.2%; Score 32; DB 21; Length 342;
XX      Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX      Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      14 GCCATCTGCAGGTCGAGTGATGATGCTG 45
XX      |||
XX      151 GCCATCTGCAGGTCGAGTGATGATGCTG 182
XX
XX      RESULT 4
XX      ID      AAH43570
XX      ID      AAH43570 standard; cDNA; 342 BP.
XX
XX      AC      AAH43570;
XX
XX      DT      07-JAN-2002 (first entry)
XX
XX      DE      ROC2 coding sequence.
XX
XX      KW      Assay; ubiquitin ligase; tag1-ubiquitin; E1; E2;
XX      KW      ubiquitin activating enzyme; ubiquitin conjugating enzyme; E3;
XX      KW      ubiquitin ligase; ubiquitination modulator; ss.
XX
XX      OS      Unidentified.
XX
XX      PN      WO200175145-A2.
XX
XX      PD      11-OCT-2001.
XX
XX      PF      03-APR-2001; 2001WO-US10906.
XX      03-APR-2000; 2000US-0542497.
XX
XX      (RIGB-) RIGEL PHARM INC.

```

```

XX      Issakant SD, Huang J, Sheung J, Pray TR;
XX      WPI; 2001-626445/72.
XX      P-PSDB; AAB47600.
XX
XX      Assaying ubiquitin ligase activity for identifying modulators of
XX      ubiquitination, by combining ubiquitin, ubiquitin activating
XX      conjugating enzyme, ubiquitin ligase and measuring amount of ubiquitin
XX      bound to the ligase
XX
XX      Example 1; Fig 12A; 98pp; English.
XX
XX      The sequences given in AAH43568-72 encode proteins which may be used
XX      in the method of the invention for assaying ubiquitin ligase activity.
XX      The method comprises combining under conditions that favour ubiquitin
XX      ligase activity, tag1-ubiquitin, E1 (ubiquitin activating enzyme), E2
XX      (ubiquitin conjugating enzyme) and E3 (ubiquitin ligase) and measuring
XX      the amount of tag1-ubiquitin bound to the E3. The method is useful for
XX      assaying ubiquitin ligase activity and ubiquitination enzyme activity
XX      which is useful for identifying ubiquitination modulator. The method
XX      comprises combining tag1-ubiquitin, the modulator, E1, E2 and tag2-E3
XX      and measuring the amount of tag1-ubiquitin bound to tag2-E3 or combining
XX      tag1-ubiquitin, modulator, E1 and tag3-E2. Ubiquitin ligase activity is measured
XX      directly, eliminating the need for target proteins and subsequent
XX      analysis such as separating ligated from unligated material in an
XX      SDS-PAGE procedure. This allows multi-well array analysis and high
XX      throughput screening techniques for modulators of ubiquitination
XX      activity. The method also allows the analysis of many different
XX      combinations of E3 components and E2/E3 combinations without requiring
XX      prior identification of specific target substrates. Ubiquitin is labeled,
XX      directly or indirectly and this allows for easy and rapid detection and
XX      measurement of ligated ubiquitin.
XX
XX      Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other;
XX
XX      Query Match      25.2%; Score 32; DB 22; Length 342;
XX      Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX      Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      14 GCCATCTGCAGGTCGAGTGATGATGCTG 45
XX      |||
XX      151 GCCATCTGCAGGTCGAGTGATGATGCTG 182
XX
XX      RESULT 5
XX      ID      AAD39674
XX      ID      AAD39674 standard; DNA; 342 BP.
XX
XX      AC      AAD39674;
XX
XX      DT      22-OCT-2002 (first entry)
XX
XX      DE      Human RING finger protein, ROC2 DNA.
XX
XX      KW      Ubiquitin ligase; U1; tag1-ubiquitin; E1; ubiquitin-activating enzyme;
XX      KW      E2; ubiquitin-conjugating enzyme/ubiquitin carrier protein; E3; human;
XX      KW      RING finger protein; gene; ds.
XX
XX      OS      Homo sapiens.
XX
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      1..342
XX      FT      /*tag= a
XX      FT      /product= "Human ROC2 protein"
XX
XX      PN      US2002042083-A1.
XX
XX      PD      11-APR-2002.
XX
XX      PF      03-APR-2001; 2001US-0826312.
XX

```

```

PR 03-APR-2000; 2000US-0542497.
XX
XX (RICE-) RIGEL PHARM INC.
XX
XX Iseakani SD, Huang J, Sheung J, Pray TR;
XX WPI, 2002-488718/52.
XX P-PSDB; AAE24621.
XX
XX Assay for ubiquitin ligase activity, useful for identifying modulators,
XX by measuring binding of labeled ubiquitin to ubiquitin ligase
XX
XX Disclosure; Fig 12A; 56pp; English.
XX
XX The invention relates to an assay for ubiquitin ligase (UL) activity
XX which comprises incubating tag1-ubiquitin, E1 (ubiquitin-activating
XX enzyme), E2 (ubiquitin-conjugating enzyme/ubiquitin carrier protein)
XX and E3 (UL) and measuring the amount of tag1-ubiquitin bound to E3. The
XX method is particularly used to screen for modulators of UL activity. The
XX present sequence is human RING finger protein, ROC2 DNA.
XX
XX Sequence 342 BP; 82 A; 84 C; 108 G; 68 T; 0 other;
XX
XX
XX Query Match 25.2%; Score 32; DB 24; Length 342;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 14 GCCATCTGCAGGCTCCAGGTGATGATGCCTG 45
XX 151 GCCATCTGCAGGCTCCAGGTGATGATGCCTG 182
XX
XX
XX RESULT 6
XX ID AAX87317 standard; cDNA; 754 BP.
XX
XX AAX87317;
XX
XX 27-SEP-1999 (first entry)
XX
XX Human sensitive to apoptosis (SAG) gene mutant MML.
XX
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX vulnerability; therapy; mutant; ds.
XX
XX Homo sapiens.
XX OS
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH 1..342
XX FT /*tag= a
XX FT replace(148,T)
XX FT mutation
XX FT /*tag= b
XX FT /note= "C50S mutation"
XX
XX WO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX P-PSDB; AAY06495.
XX
XX SAG; Sensitive to Apoptosis Gene and related proteins, useful for

```

```

PT promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 59; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MML, which codes for a SAG protein (see AAY06495)
XX in which the Cys residue at position 50 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 50 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is at a haem binding site of SAG. Single and
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. MML showed reduced haem binding but unaffected
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversal of a tumour phenotype.
XX SAG gene, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
XX
XX Query Match 25.2%; Score 32; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 14 GCCATCTGCAGGCTCCAGGTGATGATGCCTG 45
XX 151 GCCATCTGCAGGCTCCAGGTGATGATGCCTG 182
XX
XX
XX RESULT 7
XX ID AAX87321 standard; cDNA; 754 BP.
XX
XX AAX87321;
XX
XX 27-SEP-1999 (first entry)
XX
XX Human sensitive to apoptosis (SAG) gene mutant MMS.
XX
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX vulnerability; therapy; mutant; ds.
XX
XX Homo sapiens.
XX OS
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH 1..342
XX FT /*tag= a
XX FT replace(238,T)
XX FT mutation
XX FT /*tag= b
XX FT /note= "C80S mutation"
XX
XX WO9932514-A2.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX P-PSDB; AAY06499.
XX
XX

```

PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
XX  
PS Claim 15; Page 66; 84pp; English.

CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM5, which codes for a SAG protein (see AAY06499)  
CC in which the Cys residue at position 80 of the native protein (see  
CC AAY06492) is replaced by a Ser residue owing to a mutation of codon  
CC 80 from TGC to AGC obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is in zinc-binding finger 1 of SAG. Single and  
CC double SAG mutants (see AAX87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM5  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversal of a tumour phenotype.  
CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.

SQ Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;

Query Match 25.2%; Score 32; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45  
DB 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182

RESULT 8  
AAX87322  
ID AAX87322 standard; cDNA; 754 BP.  
XX  
XX AAX87322;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM6.  
XX  
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerability; therapy; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH CDS 1..342  
FT /\*tag= a  
FT replace(244..246,CAT)  
FT /\*tag= b  
FT /note= "H82K mutation"  
FT  
FT  
XX WO9932514-A2.  
XX  
XX  
XX 01-JUL-1999.  
XX  
XX  
XX 15-DEC-1998; 98WO-US26705.  
XX  
XX PF 11-SEP-1998; 98US-0099840.  
XX PR 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN ) WARNER LAMBERT CO.  
XX  
XX Sun Y;  
XX  
XX WPI; 1999-430152/36.  
XX  
XX P-PsDB; AAY06500.  
DR

XX  
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
PT  
XX  
XX  
PS Claim 15; Page 67-68; 84pp; English.

CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM6, which codes for a SAG protein (see AAY06500)  
CC in which the His residue at position 82 of the native protein (see  
CC AAY06492) is replaced by a Lys residue owing to a mutation of codon  
CC 82 from CAT to AAA obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is in zinc-binding finger 1 of SAG. Single and  
CC double SAG mutants (see AAX87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM6  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversal of a tumour phenotype.  
CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.

SQ Sequence 754 BP; 207 A; 154 C; 201 G; 192 T; 0 other;

Query Match 25.2%; Score 32; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45  
DB 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182

RESULT 9  
AAX87323  
ID AAX87323 standard; cDNA; 754 BP.  
XX  
XX AAX87323;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM7.  
XX  
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerability; therapy; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH CDS 1..342  
FT /\*tag= a  
FT replace(253..255,CAC)  
FT /\*tag= b  
FT /note= "H85K mutation"  
FT  
FT  
XX WO9932514-A2.  
XX  
XX  
XX 01-JUL-1999.  
XX  
XX  
XX 15-DEC-1998; 98WO-US26705.  
XX  
XX PF 11-SEP-1998; 98US-0099840.  
XX PR 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN ) WARNER LAMBERT CO.  
XX  
XX Sun Y;  
XX  
XX WPI; 1999-430152/36.  
XX  
XX



XX MPI; 1999-430152/36.  
DR P-PSDB; AAY06503.  
XX  
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
promoting cell growth and protecting cells against apoptosis  
XX  
PS Claim 15; Page 72-73; 84pp; English.  
XX  
CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM9, which codes for a SAG protein (see AAY06503)  
CC in which the Cys residue at position 99 of the native protein (see  
CC AAY06492) is replaced by a Ser residue owing to a mutation of codon  
CC 99 from TGC to AGC obtained by site-directed mutagenesis of SAG  
CC cDNA. This residue is in zinc-finger finger 2 of SAG. Single and  
CC double SAG mutants (see AAX87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. These properties were unaffected by the MM9  
CC mutation. SAG is a novel zinc finger protein that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.  
CC SAG genes, and mutant SAG genes, can be used to protect cells from  
CC apoptosis induced by redox reagents. They can also be used for the  
CC recombinant production of SAG proteins, which are molecular targets  
CC in the development of drugs against neurodegenerative disorders,  
CC cancers and muscle dystrophy, and promoting wound healing.  
XX  
SQ Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;  
Query Match 25.2%; Score 32; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 GCCATCTGCAGGCTCCAGTGATGATGCCTG 45  
DB 151 GCCATCTGCAGGCTCCAGTGATGATGCCTG 182  
RESULT 12  
AAX87326  
ID AAX87326 standard; cDNA; 754 BP.  
XX  
AC AAX87326;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM10.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnery; therapy; mutant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..342  
FT mutation /tag= a  
FT mutation /tag= b  
FT mutation /note= "H85K mutation"  
FT mutation replace(253..255,CAC)  
FT /tag= C  
FT /note= "H85K mutation"  
XX  
PN WO9332514-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 15-DEC-1998; 98WO-US26705.  
XX  
PR 11-SEP-1998; 98US-0099840.  
PR 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.  
PA  
XX Sun Y;  
XX  
XX MPI; 1999-430152/36.  
DR P-PSDB; AAY06504.  
XX  
PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
promoting cell growth and protecting cells against apoptosis  
XX  
PS Claim 15; Page 74-75; 84pp; English.  
XX  
CC This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene MM10, which codes for a SAG protein (see AAY06504)  
CC in which the His residues at positions 82 and 85 of the native  
CC protein (see AAY06492) are replaced by Lys residues owing to  
CC mutations of codons 82 and 85 by site-directed mutagenesis of SAG  
CC cDNA. These residues are in zinc-finger fingers 1 and 2 of SAG. SAG  
CC mutants (see AAX87317-31) were made in order to determine the role of  
CC each cysteine residue of SAG in haem binding and oligomerization.  
CC These properties were unaffected by the MM10 mutations. SAG is a  
CC novel zinc finger protein that promotes cell growth, protects cells  
CC from apoptosis, scavenges oxygen radicals and can be used for the  
CC reversion of a tumour phenotype. SAG genes, and mutant SAG genes,  
CC can be used to protect cells from apoptosis induced by redox  
CC reagents. They can also be used for the recombinant production of  
CC SAG proteins, which are molecular targets in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy, and promoting wound healing.  
XX  
SQ Sequence 754 BP; 208 A; 152 C; 202 G; 192 T; 0 other;  
Query Match 25.2%; Score 32; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 GCCATCTGCAGGCTCCAGTGATGATGCCTG 45  
DB 151 GCCATCTGCAGGCTCCAGTGATGATGCCTG 182  
RESULT 13  
AAX87327  
ID AAX87327 standard; cDNA; 754 BP.  
XX  
AC AAX87327;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant MM11.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KW neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnery; therapy; mutant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..342  
FT mutation /tag= a  
FT mutation /tag= b  
FT mutation /note= "C64S mutation"  
FT /note= "C64S mutation"  
XX  
PN WO9332514-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 15-DEC-1998; 98WO-US26705.  
XX  
PR 11-SEP-1998; 98US-0099840.

```

PR 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06505.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 76; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM11, which codes for a SAG protein (see AAY06505)
XX in which the Cys residue at position 64 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 64 from TGT to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is in zinc-binding finger 1 of SAG. Single and
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. These properties were unaffected by the MM11
XX mutation. SAG is a novel zinc finger protein that promotes
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversion of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
XX Query Match 25.2%; Score 32; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
XX |||
XX Db 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182
XX
XX RESULT 14
XX AAX87328
XX ID AAX87328 standard; cDNA; 754 BP.
XX
XX AC AAX87328;
XX
XX DT 27-SEP-1999 (first entry)
XX
XX DE Human sensitive to apoptosis (SAG) gene mutant MM12.
XX
XX KM SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX vulnerable; therapy; mutant; ds.
XX
XX OS Homo sapiens.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT 1..342
XX FT /*tag= a
XX FT replace(217,T)
XX FT /*tag= b
XX FT /note= "C73S mutation"
XX
XX PN WO9932514-A2.
XX
XX PD 01-JUL-1999.
XX
XX PF 15-DEC-1998; 98WO-US26705.
XX

```

```

PR 11-SEP-1998; 98US-0099840.
XX
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06506.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 77-78; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM12, which codes for a SAG protein (see AAY06506)
XX in which the Cys residue at position 73 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 73 from TGT to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is in a proase inhibitor motif of SAG. Single
XX and double SAG mutants (see AAX87317-31) were made to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. These properties were unaffected by the MM12
XX mutation. SAG is a novel zinc finger protein that promotes
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversion of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
XX Query Match 25.2%; Score 32; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
XX |||
XX Db 151 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 182
XX
XX RESULT 15
XX AAX87330
XX ID AAX87330 standard; cDNA; 754 BP.
XX
XX AC AAX87330;
XX
XX DT 27-SEP-1999 (first entry)
XX
XX DE Human sensitive to apoptosis (SAG) gene mutant MM14.
XX
XX KM SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX vulnerable; therapy; mutant; ds.
XX
XX OS Homo sapiens.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT 1..342
XX FT /*tag= a
XX FT replace(295,C)
XX FT /*tag= b
XX FT /note= "C99S mutation"
XX FT /note= "C304,C"
XX FT /*tag= c
XX FT /note= "C102S mutation"
XX
XX PN WO9932514-A2.
XX
XX PD 01-JUL-1999.
XX
XX PF 15-DEC-1998; 98WO-US26705.
XX

```

```

PD 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-US26705.
XX
XX 11-SEP-1998; 98US-0099840.
XX
XX 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06508.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 81; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM14, which codes for a SAG protein (see AAY06508)
XX in which the Cys residues at positions 99 and 102 of the native
XX protein (see AAY06492) are replaced by Ser residues owing to
XX mutations of codons 99 and 102 through site-directed mutagenesis of
XX SAG cDNA. These residues are in zinc ring finger 2 of SAG. Single
XX double SAG mutants (see AAX87317-31) were made in order to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. Haem binding was unaffected in MM14, but no
XX oligomerization was observed, indicating that these residues are
XX important for intermolecular disulfide bond formation. SAG is
XX a novel zinc finger protein that promotes cell growth, protects cells
XX from apoptosis, scavenges oxygen radicals and can be used for the
XX reversal of a tumour phenotype. SAG genes, and mutant SAG genes,
XX can be used to protect cells from apoptosis induced by redox
XX reagents. They can also be used for the recombinant production of
XX SAG proteins, which are molecular targets in the development of
XX drugs against neurodegenerative disorders, cancers and muscle
XX dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;
XX
XX Query Match 25.2%; Score 32; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 14 GCCATCTGCAGGTCAGGTGATGATGCCTG 45
XX |||||||||||||||||||||||||||||||
XX Db 151 GCCATCTGCAGGTCAGGTGATGATGCCTG 182
XX
XX RESULT 16
XX AAX87331
XX ID AAX87331 standard; cDNA; 754 BP.
XX
XX AC AAX87331;
XX
XX DT 27-SEP-1999 (first entry)
XX
XX DE Human sensitive to apoptosis (SAG) gene mutant MM15.
XX
XX KW SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX KW vulnerability; therapy; mutant; ds.
XX
XX OS Homo sapiens.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT 1.342
XX FT CDS /tag= a
XX FT mutation /tag= b
XX FT /note= "C47S mutation"
XX

```

```

XX
XX PN WO9932514-A2.
XX
XX XX 01-JUL-1999.
XX
XX PD 01-JUL-1999.
XX
XX PF 15-DEC-1998; 98WO-US26705.
XX
XX PF 11-SEP-1998; 98US-0099840.
XX
XX PR 19-DEC-1997; 97US-0068179.
XX
XX (WARN ) WARNER LAMBERT CO.
XX
XX Sun Y;
XX
XX WPI; 1999-430152/36.
XX
XX P-PSDB; AAY06509.
XX
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX promoting cell growth and protecting cells against apoptosis
XX
XX Claim 15; Page 82-83; 84pp; English.
XX
XX This is the nucleotide sequence of human sensitive to apoptosis
XX (SAG) mutant gene MM15, which codes for a SAG protein (see AAY06509)
XX in which the Cys residue at position 47 of the native protein (see
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon
XX 47 from TGC to AGC obtained by site-directed mutagenesis of SAG
XX cDNA. This residue is in a GADPH binding site of SAG. Single
XX and double SAG mutants (see AAX87317-31) were made to determine
XX the role of each cysteine residue of SAG in haem binding and SAG
XX oligomerization. These properties were unaffected by the MM15
XX mutation. SAG is a novel zinc finger protein that promotes
XX cell growth, protects cells from apoptosis, scavenges oxygen
XX radicals and can be used for the reversal of a tumour phenotype.
XX SAG genes, and mutant SAG genes, can be used to protect cells from
XX apoptosis induced by redox reagents. They can also be used for the
XX recombinant production of SAG proteins, which are molecular targets
XX in the development of drugs against neurodegenerative disorders,
XX cancers and muscle dystrophy, and promoting wound healing.
XX
XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;
XX
XX Query Match 25.2%; Score 32; DB 20; Length 754;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 14 GCCATCTGCAGGTCAGGTGATGATGCCTG 45
XX |||||||||||||||||||||||||||||||
XX Db 151 GCCATCTGCAGGTCAGGTGATGATGCCTG 182
XX
XX RESULT 17
XX AAX87314
XX ID AAX87314 standard; cDNA; 754 BP.
XX
XX AC AAX87314;
XX
XX DT 27-SEP-1999 (first entry)
XX
XX DE Human sensitive to apoptosis (SAG) gene.
XX
XX KW SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX neurodegenerative disease; muscular dystrophy; wound healing;
XX KW vulnerability; therapy; ds.
XX
XX OS Homo sapiens.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT 1.342
XX FT CDS /tag= a
XX FT WO9932514-A2.
XX

```

PD 01-JUL-1999.  
XX  
PF 15-DEC-1998; 98WO-US26705.  
XX  
XX 11-SEP-1998; 98US-0099840.  
PR 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN ) WARNER LAMBERT CO.  
PI Sun Y;  
DR MPI: 1999-430152/36.  
P-PSDB; AAY06492.  
XX  
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
XX  
XX Claim 9; Page 50-51; 84pp; English.  
XX  
XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) cDNA, which codes for a novel redox-sensitive, haem-binding  
CC protein (see AAY06492) with a zinc RING finger domain that promotes  
CC cell growth, protects cells from apoptosis, scavenges oxygen  
CC radicals and can be used for the reversion of a tumour phenotype.  
CC The cDNA was isolated from a HeLa cell cDNA library using mouse  
CC SAG cDNA (see AAX87313) as probe. SAG is highly conserved among  
CC species. Disruption in yeast was shown to be lethal. SAG deletion  
CC mutants (see AAX87315-16) have been identified in human cancer lines,  
CC suggesting a role in carcinogenesis. SAG genes, and mutant SAG  
CC genes, can be used to protect cells from apoptosis induced by redox  
CC reagents. Antisense SAG genes can also be used to inhibit the growth of  
CC tumour cells. The SAG proteins can also be used for the recombinant  
CC production of the SAG proteins. The SAG proteins can be used to  
CC scavenge oxygen radicals in organisms and to promote wound healing.  
CC Additionally, the SAG genes or their complements can be used to  
CC promote or inhibit the growth of plant cells (all claimed). The SAG  
CC protein is also an ideal molecular target in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy.  
XX  
SQ Sequence 754 BP; 205 A; 155 C; 201 G; 193 T; 0 other;  
XX  
Query Match 25.2%; Score 32; DB 20; Length 754;  
Best local Similarity 100.0%; Pred.No. 4.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 14 GCCATCTGCAGGGTCCAGTGTGATGATGCTG 45  
DB 151 GCCATCTGCAGGGTCCAGTGTGATGATGCTG 182  
RESULT 18  
AAS25860  
ID AAS25860 standard; cDNA; 836 BP.  
XX  
AC AAS25860;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, Seq ID 39.  
XX  
XX Human; immunosuppressive; antiarthritis; ss; antihemetic;  
KW cytoskeletal; cardiac; vasotropic; cerebroprotective; neotropic;  
KW neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX

PN WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231415.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0233223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.



PR	29-SEP-2000	2000US-0236327	
PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	
PR	02-OCT-2000	2000US-0236682	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	02-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239935	
PR	13-OCT-2000	2000US-0239937	
PR	20-OCT-2000	2000US-0240960	
PR	20-OCT-2000	2000US-0241221	
PR	20-OCT-2000	2000US-0241185	
PR	20-OCT-2000	2000US-0241787	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241809	
PR	20-OCT-2000	2000US-0241826	
PR	08-NOV-2000	2000US-0244617	
PR	08-NOV-2000	2000US-0246674	
PR	08-NOV-2000	2000US-0246675	
PR	08-NOV-2000	2000US-0246676	
PR	08-NOV-2000	2000US-0246677	
PR	08-NOV-2000	2000US-0246678	
PR	08-NOV-2000	2000US-0246679	
PR	08-NOV-2000	2000US-0246682	
PR	08-NOV-2000	2000US-0246684	
PR	08-NOV-2000	2000US-0246685	
PR	08-NOV-2000	2000US-0246686	
PR	08-NOV-2000	2000US-0246687	
PR	08-NOV-2000	2000US-0246688	
PR	08-NOV-2000	2000US-0246689	
PR	08-NOV-2000	2000US-0246690	
PR	08-NOV-2000	2000US-0246691	
PR	08-NOV-2000	2000US-0246692	
PR	08-NOV-2000	2000US-0246693	
PR	08-NOV-2000	2000US-0246694	
PR	08-NOV-2000	2000US-0246695	
PR	08-NOV-2000	2000US-0246696	
PR	08-NOV-2000	2000US-0246697	
PR	08-NOV-2000	2000US-0246698	
PR	08-NOV-2000	2000US-0246699	
PR	08-NOV-2000	2000US-0246700	
PR	08-NOV-2000	2000US-0246701	
PR	08-NOV-2000	2000US-0246702	
PR	08-NOV-2000	2000US-0246703	
PR	08-NOV-2000	2000US-0246704	
PR	08-NOV-2000	2000US-0246705	
PR	08-NOV-2000	2000US-0246706	
PR	08-NOV-2000	2000US-0246707	
PR	08-NOV-2000	2000US-0246708	
PR	08-NOV-2000	2000US-0246709	
PR	08-NOV-2000	2000US-0246710	
PR	08-NOV-2000	2000US-0246711	
PR	08-NOV-2000	2000US-0246712	
PR	08-NOV-2000	2000US-0246713	
PR	08-NOV-2000	2000US-0246714	
PR	08-NOV-2000	2000US-0246715	
PR	08-NOV-2000	2000US-0246716	
PR	08-NOV-2000	2000US-0246717	
PR	08-NOV-2000	2000US-0246718	
PR	08-NOV-2000	2000US-0246719	
PR	08-NOV-2000	2000US-0246720	
PR	08-NOV-2000	2000US-0246721	
PR	08-NOV-2000	2000US-0246722	
PR	08-NOV-2000	2000US-0246723	
PR	08-NOV-2000	2000US-0246724	
PR	08-NOV-2000	2000US-0246725	
PR	08-NOV-2000	2000US-0246726	
PR	08-NOV-2000	2000US-0246727	
PR	08-NOV-2000	2000US-0246728	
PR	08-NOV-2000	2000US-0246729	
PR	08-NOV-2000	2000US-0246730	
PR	08-NOV-2000	2000US-0246731	
PR	08-NOV-2000	2000US-0246732	
PR	08-NOV-2000	2000US-0246733	
PR	08-NOV-2000	2000US-0246734	
PR	08-NOV-2000	2000US-0246735	
PR	08-NOV-2000	2000US-0246736	
PR	08-NOV-2000	2000US-0246737	
PR	08-NOV-2000	2000US-0246738	
PR	08-NOV-2000	2000US-0246739	
PR	08-NOV-2000	2000US-0246740	

XX	WPI; 2001-468783/53.
DR	P-PSDB; AAU15873.
XX	New nucleic acid molecules encoding 461 human secreted proteins for
PT	diagnosing, preventing, treating or ameliorating medical conditions and
PT	used as food additives or preservatives -
XX	
PS	Claim 1; SEQ ID No 39; 980bp; English.
XX	
CC	The invention relates to isolated nucleic acid molecules and their
CC	encoded secreted proteins. The nucleic acids and proteins are used to
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC	are also used in diagnosing a pathological condition or susceptibility
CC	to a pathological condition. Antibodies to the proteins can also
CC	be used in alleviating symptoms associated with the disorders and in
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC	immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC	include autoimmune diseases e.g. rheumatoid arthritis,
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC	e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC	and ocular disorders e.g. corneal infection, and many other
CC	disorders listed in the specification. The polypeptides can also
CC	be used to aid wound healing and epithelial cell proliferation, to
CC	prevent skin aging due to sunburn, to maintain organs before
CC	transplantation, for supporting cell culture of primary tissues, to
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used
CC	as a food additive or preservative to increase or decrease storage
CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC	minerals, cofactors and other nutritional components. The present
CC	sequence encodes a novel secreted protein of the invention.
OY	Query Match 25.2%; Score 32; DB 22; Length 836;
	Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Dn	Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
14	GCCATCTGCAGGGTCGATGATGATGCCCTG 45
166	GCCATCTGCAGGGTCGATGATGATGCCCTG 197
RESULT 19	
AAS26319	ID AAS26319 standard; cDNA; 836 BP.
XX	AAS26319;
AC	
DT	07-NOV-2001 (first entry)
XX	
DE	Human cDNA encoding a novel secreted protein, Seq ID 498.
KM	Human; immunosuppressive; antiarthritis; ss; antirheumatic;
XX	Cytostatic; cardiact; vasotropic; cerebroprotective; nootropic;
XX	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KM	vulnerable; secreted protein; rheumatoid arthritis;
XX	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KM	cerebrovascular disorder; cerebral ischemia; angiogenesis;
KM	nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KM	corneal infection; wound healing; epithelial cell proliferation;
XX	skin ageing; food additive; preservative; antiproliferative.
OS	Homo sapiens.
XX	
XX	WO200155322-A2.
PN	
XX	02-AUG-2001.
PD	
PF	17-JAN-2001; 2001MO-US01341.
RR	31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 12-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0234222.  
PR 21-SEP-2000; 2000US-0234227.  
PR 21-SEP-2000; 2000US-0234227.  
PR 21-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239395.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249254.  
PR 17-NOV-2000; 2000US-0249255.  
PR 17-NOV-2000; 2000US-0249257.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251988.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-488783/53.  
XX P-PSDB; AAUI6332.  
XX  
PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

XX Claim 1; SEQ ID No 498; 980bp; English.  
PS  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence encodes a novel secreted protein of the invention.

Query Match 25.2%; Score 32; DB 22; Length 836;  
Best Local Similarity 100.0%; Pred. No. 4,5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GCCATCTGCAGGCTCCAGTGTATGATGCCTG 45  
|||  
Db 219 GCCATCTGCAGGCTCCAGTGTATGATGCCTG 250

RESULT 20  
ABX73201  
ID ABX73201 standard; DNA; 836 BP.  
XX  
AC ABX73201;  
XX  
DT 18-MAR-2003 (first entry)  
XX  
DE Human novel polynucleotide #29.  
XX  
KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;  
KW muscular disorder; respiratory disease; reproductive disorder;  
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
KW cardiovascular; nephrotoxic; cytostatic; antiallergic; thrombolytic;  
KW haemostatic; antiarteriosclerotic.  
XX  
XX Homo sapiens.  
XX  
XX US2002132753-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 17-JAN-2001; 2001US-0764864.  
XX  
PR 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 28-JUN-2000; 2000US-214886P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.

PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 22-AUG-2000; 2000US-226688P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 08-DEC-2000; 2000US-251856P.  
PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251869P.

(ROSE/) ROSEN C A.  
(RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI; 2003-147444/14.  
XX P-PSDB; ABUS4941.  
XX  
XX  
PT New polypeptides and nucleic acids, useful in gene therapy for  
PT treating, inhibiting or preventing e.g. neural, immune system,  
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,  
PT cardiovascular or renal disorders -  
XX  
XX  
PS Claim 1; SEQ ID NO 39; 402pp; English.  
XX  
XX  
CC The invention relates to human novel polypeptides and their associated  
CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid  
CC arthritis and multiple sclerosis), muscular disorders, respiratory  
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),  
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,  
CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's  
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute  
CC kidney failure and end-stage renal disease), hyperproliferative disorders  
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.  
CC septic shock, bursitis and appendicitis), allergic reactions and  
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,  
CC atherosclerosis and myocardial infarction) and cancerous diseases.  
CC Sequences ABX7313-ABX74167 represent human novel polynucleotides of the  
CC invention.

```

XX SQ Sequence 836 BP; 264 A; 168 C; 209 G; 195 T; 0 other;
Query Match 25.2%; Score 32; DB 25; Length 836;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 14 GCCATCTGCAGGTCACAGTGTGATGATGCTTG 45
Db 166 GCCATCTGCAGGTCACAGTGTGATGATGCTTG 197

RESULT 21
ABX73660
ID ABX73660 standard; DNA; 836 BP.
XX AC ABX73660;
XX DT 18-MAR-2003 (first entry)
XX DE Human novel polynucleotide #488.
XX KW Human; gene; dg; neural disorder; immune system disorder; renal disorder;
muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotoxic; cytotoxic; antiallergic; thrombolytic;
KW haemostatic; antiterosclerotic.
XX OS Homo sapiens.
XX PN US2002132753-A1.
XX PD 19-SEP-2002.
XX PF 17-JUN-2001; 2001US-0764864.
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 28-JUN-2000; 2000US-214886P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-217496P.
XX 14-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 22-AUG-2000; 2000US-226868P.
XX 30-AUG-2000; 2000US-226924P.
XX 01-SEP-2000; 2000US-229287P.
XX 01-SEP-2000; 2000US-229343P.
XX 01-SEP-2000; 2000US-229344P.
XX 01-SEP-2000; 2000US-229345P.
XX 05-SEP-2000; 2000US-229509P.
XX 05-SEP-2000; 2000US-229513P.
XX 08-SEP-2000; 2000US-231413P.
XX 21-SEP-2000; 2000US-234223P.
XX 21-SEP-2000; 2000US-234274P.
XX 25-SEP-2000; 2000US-234979P.
XX 27-SEP-2000; 2000US-235834P.
XX 29-SEP-2000; 2000US-236327P.
XX 29-SEP-2000; 2000US-236367P.
XX 29-SEP-2000; 2000US-236368P.
XX 29-SEP-2000; 2000US-236369P.

```

```

PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX (ROSE/) ROSEN C A.
XX (ROSE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX Rosen CA, Ruben SM, Barash SC;
XX MPI; 2003-147444/14.
XX P-PSDB; ABUS5400.
DR New polypeptides and nucleic acids, useful in gene therapy for
PT treating, inhibiting or preventing e.g. neural, immune system,
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT cardiovascular or renal disorders -
XX Claim 1; SEQ ID NO 498; 402pp; English.
XX
XX CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC arthritis and multiple sclerosis), muscular disorders, respiratory
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC cardiovascular disorders (e.g. congenital heart defects, Epstein's
CC anomaly and hypoplastic left heart disease), hyperproliferative disorders
CC kidney failure and end-stage renal disease), inflammatory disorders
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC septic shock, bursitis and appendicitis), allergic reactions and
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC atherosclerosis and myocardial infarction) and cancerous diseases.
CC Sequences ABX73173-ABX74167 represent human novel polynucleotides of the
CC invention.
XX
XX SQ Sequence 836 BP; 224 A; 179 C; 221 G; 211 T; 1 other;
Query Match 25.2%; Score 32; DB 25; Length 836;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 14 GCCATCTGCAGGTCACAGTGTGATGATGCTTG 45
Db 219 GCCATCTGCAGGTCACAGTGTGATGATGCTTG 250

RESULT 22
AAC77504
ID AAC77504 standard; cDNA; 962 BP.
XX AAC77504;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF3059 polynucleotide sequence SEQ ID NO.6117.
XX Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;
KW vulnertary; antipariotatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antitartritic; immunosuppressant; cardiant;
KW

```

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 OS Homo sapiens.  
 XX WO200058473-A2.  
 PN 05-OCT-2000.  
 XX 31-MAR-2000; 2000WO-US08621.  
 PF 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX (CURA-) CURAGEN CORP.  
 PA Shinkens RA, Leach M;  
 PI WPI; 2000-602362/57.  
 DR P-PSDB; AAB433295.  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PT Claim 5; Page 5300; 5507pp; English.  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
 CC osteoparalytic; anticonvulsant; antiallergic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antirheumatic; antiviral; antifungal; antineoplastic;  
 CC antihypertensive; antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineoplastic disease, to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC XX  
 SQ Sequence 962 BP; 265 A; 204 C; 243 G; 250 T; 0 other;

Query Match 25.2%; Score 32; DB 21; Length 962;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCAGGTCAGTGCCTG 45  
 DB 202 GCCATCTGCAGGTCAGGTCAGTGCCTG 233

RESULT 23  
 AAX87320

ID AAX87320 standard; cDNA; 754 BP.  
 XX AAX87320;  
 AC 27-SEP-1999 (first entry)  
 DT Human sensitive to apoptosis (SAG) gene mutant MM4.  
 DE SAG gene; sensitive to apoptosis; human; cancer; tumour;  
 KW neurodegenerative disease; muscular dystrophy; wound healing;  
 KW vulnary; therapy; mutant; ds.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 XX CDS 1..342  
 FT /\*tag= a  
 FT mutation replace(181,T)  
 FT /\*tag= b  
 FT /note= "C61S mutation"  
 XX WO9932514-A2.  
 PN 01-JUL-1999.  
 PD 15-DEC-1998; 98WO-US26705.  
 PF 11-SEP-1998; 98US-0099840.  
 PR 19-DEC-1997; 97US-0068179.  
 PR (WARN ) WARNER LAMBERT CO.  
 PA Sun Y;  
 PI WPI; 1999-430152/36.  
 DR P-PSDB; AAY06498.  
 XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
 PT promoting cell growth and protecting cells against apoptosis  
 PT Claim 15; Page 64-65; 84pp; English.  
 XX This is the nucleotide sequence of human sensitive to apoptosis  
 CC (SAG) mutant gene MM4, which codes for a SAG protein (see AAY06498)  
 CC in which the Cys residue at position 61 of the native protein (see  
 CC AAY06492) is replaced by a Ser residue owing to a mutation of codon  
 CC 50 from TGC to AGC obtained by site-directed mutagenesis of SAG  
 CC cDNA. This residue is at zinc-finger finger 1 of SAG. Single and  
 CC double SAG mutants (see AAX87317-11) were made in order to determine  
 CC the role of each cysteine residue of SAG in haem binding and SAG  
 CC oligomerization. These properties were unaffected by the MM4  
 CC mutation. SAG is a novel zinc finger protein that promotes  
 CC cell growth, protects cells from apoptosis, scavenges oxygen  
 CC radicals and can be used for the reversion of a tumour phenotype.  
 CC SAG genes, and mutant SAG genes, can be used to protect cells from  
 CC apoptosis induced by redox reagents. They can also be used for the  
 CC recombinant production of SAG proteins, which are molecular targets  
 CC in the development of drugs against neurodegenerative disorders,  
 CC cancers and muscle dystrophy, and promoting wound healing.  
 CC XX  
 SQ Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;

Query Match 23.6%; Score 30; DB 20; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-06;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGTCAGGTCAGTGCCTG 43  
 DB 151 GCCATCTGCAGGTCAGGTCAGTGCCTG 180

RESULT 24

AAx87329  
ID AAX87329 standard; cDNA; 754 BP.  
AC AAX87329;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant M13.  
XX  
KW SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KM neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerability; therapy; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
OS  
FH Key Location/Qualifiers  
FT CDS 1..342  
FT FT /\*tag= a  
FT FT replace(181,T)  
FT FT /\*tag= b  
FT FT /note= "C61S mutation"  
FT FT replace(190,T)  
FT FT /\*tag= c  
FT FT /note= "C64S mutation"  
XX  
XX WO9932514-A2.  
XX  
XX 01-JUL-1999.  
XX  
XX 15-DEC-1998; 98WO-US26705.  
XX  
XX 11-SEP-1998; 98US-0099840.  
XX 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN) WARNER LAMBERT CO.  
XX  
XX Sun Y;  
XX WPI; 1999-430152/36.  
XX P-PSDB; AAY06507.  
XX  
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
PT  
XX  
XX Claim 15; Page 79-80; 84pp; English.  
XX  
XX This is the nucleotide sequence of human sensitive to apoptosis  
CC (SAG) mutant gene M13, which codes for a SAG protein (see AAY06507)  
CC in which the Cys residues at positions 61 and 64 of the native  
CC mutations (see AAY06492) are replaced by Ser residues owing to  
CC mutations of codons 61 and 64 through site-directed mutagenesis of  
CC SAG cDNA. These residues are in zinc ring finger 1 of SAG. Single  
CC double SAG mutants (see AAX87317-31) were made in order to determine  
CC the role of each cysteine residue of SAG in haem binding and SAG  
CC oligomerization. M13 showed greatly reduced haem binding (and a  
CC decreased brown colour) but oligomerization was unaffected. SAG is  
CC a novel zinc finger protein that promotes cell growth, protects cells  
CC from apoptosis, scavenges oxygen radicals and can be used for the  
CC reversion of a tumour phenotype. SAG genes, and mutant SAG genes,  
CC can be used to protect cells from apoptosis induced by redox  
CC reagents. They can also be used for the recombinant production of  
CC SAG proteins, which are molecular targets in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy, and promoting wound healing.  
XX  
XX Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;  
SQ

Query Match 23.6%; Score 30; DB 20; Length 754;  
Best Local Similarity 100.0%; Pred. No. 5.2e-06;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

14 GCCATCTGCAGGCTCAGGTGATGATGCC 43

Db  
151 GCCATCTGCAGGCTCAGGTGATGATGCC 180  
|||||  
RESULT 25  
AAX87315  
ID AAX87315 standard; cDNA; 747 BP.  
XX  
XX AAX87315;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human sensitive to apoptosis (SAG) gene mutant 1.  
XX  
XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
KM neurodegenerative disease; muscular dystrophy; wound healing;  
XX vulnerability; therapy; mutant; colon carcinoma; ds.  
XX  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT CDS 1..273  
FT FT /\*tag= a  
XX  
XX WO9932514-A2.  
XX  
XX 01-JUL-1999.  
XX  
XX 15-DEC-1998; 98WO-US26705.  
XX  
XX 11-SEP-1998; 98US-0099840.  
XX 19-DEC-1997; 97US-0068179.  
XX  
XX (WARN) WARNER LAMBERT CO.  
XX  
XX Sun Y;  
XX WPI; 1999-430152/36.  
XX P-PSDB; AAY06493.  
XX  
XX SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
PT promoting cell growth and protecting cells against apoptosis  
PT  
XX  
XX Claim 15; Page 54-55; 84pp; English.  
XX  
XX This is the nucleotide sequence of deletion mutant 1 of the novel  
CC human sensitive to apoptosis gene SAG (see also AAX87314). The cDNA  
CC was obtained by PCR amplification of RNA isolated from DLD-1 colon  
CC carcinoma cells (ATCC CCL 221). It contains a 7 bp deletion  
CC compared with native SAG, starting at nucleotide 170 of the coding  
CC region that codes for a potential haem binding site. The frame  
CC shift deletion abolishes the downstream zinc finger in the  
CC resulting protein (see AAY06493). The mutation was detected by PCR  
CC in SAG RNA, but not in genomic DNA. It suggests a possible role  
CC for SAG in human carcinogenesis.  
CC genes, can be used to protect cells from apoptosis induced by redox  
CC reagents. Antisense SAG genes can also be used to inhibit the growth of  
CC tumour cells. The SAG genes can also be used for the recombinant  
CC production of SAG proteins. The SAG proteins can be used to  
CC scavenge oxygen radicals in organisms and to promote wound healing.  
CC Additionally, the SAG genes or their complements can be used to  
CC promote or inhibit the growth of plant cells (all claimed). The SAG  
CC protein is also an ideal molecular target in the development of  
CC drugs against neurodegenerative disorders, cancers and muscle  
CC dystrophy.  
XX  
XX Sequence 747 BP; 204 A; 155 C; 197 G; 191 T; 0 other;  
SQ

Query Match 22.8%; Score 29; DB 20; Length 747;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

71 CAAGAGACTGTGTGTGTGCTGGGAGA 99

Db 201 CAAGAGACTGTGTGCTGGGAGA 229

## RESULT 26

AAx87318  
ID AAX87318 standard; cDNA; 754 BP.

AC AAX87318;

DT 27-SEP-1999 (first entry)

DE Human sensitive to apoptosis (SAG) gene mutant MM2.

XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnerability; therapy; mutant; ds.

XX Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..342  
FT /\*tag= a  
FT replace(157,T)  
FT /\*tag= b  
FT /note= "C53S mutation"

XX MO9932514-A2.

XX 01-JUL-1999.

XX 15-DEC-1998; 98WO-US26705.

XX 11-SEP-1998; 98US-0099840.  
XX 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

XX Sun Y;

XX WPI; 1999-430152/36.  
XX P-PSDB; AAY06496.

PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
promoting cell growth and protecting cells against apoptosis

XX Claim 15; Page 60-61; 84pp; English.

XX This is the nucleotide sequence of human sensitive to apoptosis  
XX (SAG) mutant gene MM2, which codes for a SAG protein (see AAY06496)  
XX in which the Cys residue at position 53 of the native protein (see  
XX AAY06492) is replaced by a Ser residue owing to a mutation of codon  
XX 50 from TGC to AGC obtained by site-directed mutagenesis of SAG  
XX cDNA. This residue is at a haem binding site of SAG. Single and  
XX double SAG mutants (see AAX87317-31) were made in order to determine  
XX the role of each cysteine residue of SAG in haem binding and SAG  
XX oligomerization. These properties were unaffected by the MM2  
XX mutation. SAG is a novel zinc finger protein that promotes  
XX cell growth, protects cells from apoptosis, scavenges oxygen  
XX radicals and can be used for the reversal of a tumour phenotype.  
XX SAG genes, and mutant SAG genes, can be used to protect cells from  
XX apoptosis induced by redox reagents. They can also be used for the  
XX recombinant production of SAG proteins, which are molecular targets  
XX in the development of drugs against neurodegenerative disorders,  
XX cancers and muscle dystrophy, and promoting wound healing.

XX Sequence 754 BP; 206 A; 155 C; 201 G; 192 T; 0 other;

XX Query Match 22.8%; Score 29; DB 20; Length 754;  
XX Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 CAAGAGACTGTGTGCTGGGAGA 99  
Db 208 CAAGAGACTGTGTGCTGGGAGA 236

## RESULT 27

AAx87319  
ID AAX87319 standard; cDNA; 754 BP.

AC AAX87319;

DT 27-SEP-1999 (first entry)

DE Human sensitive to apoptosis (SAG) gene mutant MM3.

XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
XX neurodegenerative disease; muscular dystrophy; wound healing;  
KW vulnerability; therapy; mutant; ds.

XX Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..342  
FT /\*tag= a  
FT replace(148,T)  
FT /\*tag= b  
FT /note= "C50S mutation"  
FT replace(157,T)  
FT /\*tag= c  
FT /note= "C53S mutation"

XX MO9932514-A2.

XX 01-JUL-1999.

XX 15-DEC-1998; 98WO-US26705.

XX 11-SEP-1998; 98US-0099840.  
XX 19-DEC-1997; 97US-0068179.

XX (WARN ) WARNER LAMBERT CO.

XX Sun Y;

XX WPI; 1999-430152/36.  
XX P-PSDB; AAY06497.

PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
promoting cell growth and protecting cells against apoptosis

XX Claim 15; Page 62-63; 84pp; English.

XX This is the nucleotide sequence of human sensitive to apoptosis  
XX (SAG) mutant gene MM3, which codes for a SAG protein (see AAY06497)  
XX in which the Cys residues at positions 50 and 53 of the native  
XX protein (see AAY06492) are replaced by Ser residues owing to  
XX site-directed mutagenesis of SAG cDNA. These residues are at a haem  
XX binding site of SAG. Single and double SAG mutants (see AAX87317-31)  
XX were made in order to determine the role of each cysteine residue  
XX of SAG in haem binding and SAG oligomerization. MM3 showed greatly  
XX reduced haem binding but oligomerization was unaffected. SAG is a  
XX novel zinc finger protein that promotes cell growth, protects cells  
XX from apoptosis, scavenges oxygen radicals and can be used for the  
XX reversal of a tumour phenotype. SAG genes, and mutant SAG genes,  
XX can be used to protect cells from apoptosis induced by redox  
XX reagents. They can also be used for the recombinant production of  
XX SAG proteins, which are molecular targets in the development of  
XX drugs against neurodegenerative disorders, cancers and muscle  
XX dystrophy, and promoting wound healing.

XX Sequence 754 BP; 207 A; 155 C; 201 G; 191 T; 0 other;

Query Match 22.8%; Score 29; DB 20; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGTGTGGGAGA 99  
 |||||  
 DB 208 CAAGAGACTGTGTGTGTGGGAGA 236

RESULT 28  
 AAH25847  
 ID AAH25847 standard; DNA; 1152 BP.

XX AAH25847;

DT 20-AUG-2001 (first entry)

XX Human apoptosis associated protein 12 coding sequence.

XX Human; apoptosis associated protein 12; SAG protein 12; cancer;  
 KM haemopathy; HIV infection; immunological disease; inflammation; ds.

XX Homo sapiens.

OS WO200132863-A1.

PN 10-MAY-2001.

XX 30-OCT-2000; 2000MO-CN00406.

XX 29-OCT-1999; 99CN-0119924.

XX (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.

XX Mao Y, Xie Y;

XX WPI; 2001-335831/35.

XX P-PSDB; AAB98975.

XX Human apoptosis associated protein 12 and encoded polynucleotide,  
 PT applicable in diagnosis and treatment of malignant tumour, hemopathy,  
 PT HIV infection, immunological diseases and various inflammation

XX Claim 6; Page 22-23; 27pp; Chinese.

XX The present invention provides the protein and coding sequences of human  
 CC apoptosis associated (SAG) protein 12. These sequences can be used in  
 CC the diagnosis and treatment of malignant tumours, haemopathy, HIV  
 CC infection, immunological diseases and various types of inflammation. The  
 CC present sequence is the SAG protein 12 coding sequence.

XX Sequence 1152 BP; 292 A; 265 C; 309 G; 286 T; 0 other;

Query Match 22.8%; Score 29; DB 22; Length 1152;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGTGTGGGAGA 99  
 |||||  
 DB 596 CAAGAGACTGTGTGTGTGGGAGA 624

RESULT 29  
 ABN40537  
 ID ABN40537 standard; DNA; 60 BP.

XX ABN40537;

DT 15-JUN-2002 (first entry)

XX Human spliced transcript detection oligonucleotide SEQ ID NO:13285.

KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KM splice variant; transcriptome; oligonucleotide library; ss.

XX Homo sapiens.

XX WO200210449-A2.

XX 20-JUL-2001; 2001WO-IB01903.

XX 28-JUL-2000; 2000US-221607P.

XX 02-MAY-2001; 2001US-287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI; 2002-257363/30.

XX New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 PT developmental-specific genes

PS Example 1; SEQ ID 13285; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridizing selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterizing the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 BP; 10 A; 13 C; 25 G; 12 T; 0 other;

Query Match 19.7%; Score 25; DB 24; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGGTCCAGGTGATG 38  
 |||||  
 DB 36 GCCATCTGCAGGGTCCAGGTGATG 60

RESULT 30  
 ABZ11414  
 ID ABZ11414 standard; cDNA; 439 BP.

XX ABZ11414;

DT 20-JAN-2003 (first entry)

XX Human polynucleotide SEQ ID NO 296.



KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antithrombotic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US05095.  
XX  
PR 05-MAR-2001; 2001US-0799451.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Zhou P, Goodrich RW, Adundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
DR WPI; 2002-759812/82.  
XX  
P-PSDB; ABPE9197.  
XX  
PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders -  
XX  
PS Claim 1; SEQ ID NO 296; 1012bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (AB21119-AB212066) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.  
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burns, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 439 BP; 69 A; 142 C; 162 G; 66 T; 0 other;  
XX  
Query Match 19.7%; Score 25; DB 24; Length 439;  
Best Local Similarity 100.0%; Pred. No. 0.0023;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 14 GCCATCTGCAGGCTCCAGGTGATG 38  
Db 231 GCCATCTGCAGGCTCCAGGTGATG 255  
XX  
RESULT 31  
AACT7493  
XX AACT7493 standard; cDNA; 441 BP.  
XX  
AC AACT7493;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF3048 polynucleotide sequence SEQ ID NO:6095.

XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiparotitic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antithrombotic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
KW antiviral; antibacterial; antifungal; antineumatic; antihydro;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
XX  
PR 02-APR-1999; 99US-0127636.  
XX  
PR 05-APR-1999; 99US-0127728.  
XX  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
XX  
P-PSDB; AAB43284.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 5; Page 5279; 5507bp; English.  
XX  
XX AACT7446 to AACT7706 encode the proteins given in AAB40237 to AAB43397,  
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;  
XX antiparotitic; antiparkinsonian; nootropic; neuroprotective;  
XX osteopathic; anticonvulsant; antithrombotic; immunosuppressant;  
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
XX antidiabetic; hypotensive; dermatological; immunosuppressive;  
XX antineoplastic; antineumatic; antiviral; antifungal; antineumatic;  
XX antihydro; and antianemic. The sequences can be used for determining  
XX the presence of or predisposition to, or preventing or treating  
XX pathological conditions associated with an ORFX-associated disorder. The  
XX nucleic acids can be used to express ORFX proteins in gene therapy  
XX vectors. The proteins and nucleic acids may be used to treat cancers,  
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,  
XX graft vs host disease, cardiovascular disease, diabetes mellitus,  
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
XX coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 441 BP; 98 A; 113 C; 139 G; 90 T; 1 other;  
XX  
Query Match 19.7%; Score 25; DB 21; Length 441;  
Best Local Similarity 100.0%; Pred. No. 0.0023;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 14 GCCATCTGCAGGCTCCAGGTGATG 38  
Db 184 GCCATCTGCAGGCTCCAGGTGATG 208

```

RESULT 32
AA87316
ID   AA87316 standard; cDNA; 706 BP.
XX
AC   AA87316;
XX
DT   27-SEP-1999 (first entry)
XX
DE   Human sensitive to apoptosis (SAG) gene mutant 2.
XX
KW   SAG gene; sensitive to apoptosis; human; cancer; tumour;
XX   neurodegenerative disease; muscular dystrophy; wound healing;
XX   vulnery; therapy; mutant; colon carcinoma; ds.
XX
OS   Homo sapiens.
XX
FH   Key Location/Qualifiers
FT   CDS 1..294
          /*tag= a
XX
PN   MO9932514-A2.
XX
PD   01-JUL-1999.
XX
PF   15-DEC-1998; 98WO-US26705.
XX
PR   11-SEP-1998; 98US-009840.
XX   19-DEC-1997; 97US-0068179.
XX
PA   (WARN ) WARNER LAMBERT CO.
XX
PI   Sun Y;
XX
DR   WPI; 1999-430152/36.
XX   P-PSDB; AA06494.
XX
PT   SAG: Sensitive to Apoptosis Gene and related proteins, useful for
XX   promoting cell growth and protecting cells against apoptosis
XX
PS   Claim 15; Page 55-56; 84pp; English.
XX
CC   This is the nucleotide sequence of deletion mutant 2 of the novel
CC   human sensitive to apoptosis gene SAG (see also AA87314). The cDNA
CC   was obtained by PCR amplification of RNA isolated from DLD-1 colon
CC   carcinoma cells (ATCC CCL 221). It contains a 48 bp deletion
CC   compared with native SAG, starting at nucleotide 177 of the coding
CC   region that codes for a potential haem binding site. The in-frame
CC   deletion eliminates 16 amino acids in the encoded protein (see
CC   AA06494) but retains the zinc finger motif. The mutation was
CC   detected by PCR in SAG RNA, but not in genomic DNA. It suggests a
CC   possible role for SAG in human carcinogenesis. SAG genes, and
CC   mutant SAG genes, can be used to protect cells from apoptosis
CC   induced by redox reagents. Antisense SAG genes can be used to
CC   inhibit the growth of tumour cells. The SAG genes can also be used
CC   for the recombinant production of the SAG proteins. The SAG
CC   proteins can be used to scavenge oxygen radicals in organisms and
CC   to promote wound healing. Additionally, the SAG genes or their
CC   complements can be used to promote or inhibit the growth of plant
CC   cells (all claimed). The SAG protein is also an ideal molecular
CC   target in the development of drugs against neurodegenerative
CC   disorders, cancers and muscle dystrophy.
XX
SQ   Sequence 706 BP; 189 A; 147 C; 189 G; 181 T; 0 other;
XX
Query Match 19.7%; Score 25; DB 20; Length 706;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GCCATCTGCAGGGTCCAGGTGATGG 38
DB 151 GCCATCTGCAGGGTCCAGGTGATGG 175

```

```

RESULT 33
AAL16209
ID   AAL16209 standard; cDNA; 264 BP.
XX
AC   AAL16209;
XX
DT   07-DEC-2001 (first entry)
XX
DE   Human breast cancer expressed polynucleotide 8666.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
XX   14-MAR-2000; 2000US-0189167.
XX   24-MAR-2000; 2000US-0192099.
XX   29-MAR-2000; 2000US-0193480.
XX   15-MAY-2000; 2000US-0205230.
XX   09-JUN-2000; 2000US-0211315.
XX   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 1568; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterizing treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 264 BP; 64 A; 59 C; 74 G; 62 T; 5 other;
XX
Query Match 16.5%; Score 21; DB 22; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 107 CATTCCTCCACAACACTGCTGC 127
DB 57 CATTCCTCCACAACACTGCTGC 77
XX
RESULT 34
AAL25052
ID   AAL25052 standard; cDNA; 596 BP.
XX
AC   AAL25052;
XX
DT   07-DEC-2001 (first entry)
XX
DE   Human breast cancer expressed polynucleotide 17509.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX

```

OS Homo sapiens.  
XX  
PN WO200151628-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 10-JAN-2001; 2001WO-US00798.  
XX  
PR 14-JAN-2000; 2000US-0176077.  
PR 14-MAR-2000; 2000US-0189167.  
PR 24-MAR-2000; 2000US-0192099.  
PR 29-MAR-2000; 2000US-0193480.  
PR 15-MAY-2000; 2000US-0205230.  
PR 09-JUN-2000; 2000US-0211315.  
PR 25-JUL-2000; 2000US-0220534.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PI Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
DR WPI; 2001-451856/48.  
XX  
PT New peptide useful as a marker for the diagnosis of breast cancer -  
XX  
PS Claim 1; Page 3235; 3695pp; English.  
XX  
CC The invention relates to human breast cancer expressed polynucleotides  
CC (AA07544-AA026789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterising treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity.  
XX  
SQ Sequence 596 BP; 166 A; 124 C; 144 G; 162 T; 0 other;  
XX  
QY Query Match 16.5%; Score 21; DB 22; Length 596;  
Best Local Similarity 100.0%; Pred.No. 0.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
Db 107 CATTCCTCCACAACTGCTGC 127  
14 CATTCCTCCACAACTGCTGC 34  
XX  
RESULT 35  
AAL22239  
ID AAL22239 standard; cDNA; 630 BP.  
XX  
AC AAL22239;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human breast cancer expressed polynucleotide 14696.  
XX  
KW Human; breast cancer; cell marker; cytostatic; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151628-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 10-JAN-2001; 2001WO-US00798.  
XX  
PR 14-JAN-2000; 2000US-0176077.  
PR 14-MAR-2000; 2000US-0189167.  
PR 24-MAR-2000; 2000US-0192099.  
PR 29-MAR-2000; 2000US-0193480.  
PR 15-MAY-2000; 2000US-0205230.  
PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
PI Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
DR WPI; 2001-451856/48.  
XX  
PT New peptide useful as a marker for the diagnosis of breast cancer -  
XX  
PS Claim 1; Page 2649-2650; 3695pp; English.  
XX  
CC The invention relates to human breast cancer expressed polynucleotides  
CC (AA07544-AA026789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterising treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity.  
XX  
SQ Sequence 630 BP; 213 A; 125 C; 121 G; 171 T; 0 other;  
XX  
QY Query Match 16.5%; Score 21; DB 22; Length 630;  
Best Local Similarity 100.0%; Pred.No. 0.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
Db 107 CATTCCTCCACAACTGCTGC 127  
46 CATTCCTCCACAACTGCTGC 66  
XX  
RESULT 36  
ABQ90662  
ID ABQ90662 standard; DNA; 225 BP.  
XX  
AC ABQ90662;  
XX  
DT 01-OCT-2002 (first entry)  
XX  
DE M. capsulatus gene #647 for DNA array.  
XX  
KW Micro array; gene; ds; differential expression; gene expression.  
XX  
OS Methylococcus capsulatus.  
XX  
PN WO200255655-A2.  
XX  
PD 18-JUL-2002.  
XX  
PF 14-JAN-2002; 2002WO-NO00019.  
XX  
PR 12-JAN-2001; 2001NO-0000235.  
PR 12-JAN-2001; 2001NO-0000239.  
XX  
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
PA (TIGR-) TIGR.  
XX  
PI Birkefeld NK, Eichhammer I, Jonassen I, Jensen HB, Lien T;  
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;  
PI Salzberg SL;  
XX  
DR WPI; 2002-557818/59.  
XX  
PT Novel DNA array useful for determining differential expression of  
PT Methylococcus capsulatus genes, comprises polynucleotides or  
PT oligonucleotides representative for a selective number of Methylococcus  
PT capsulatus genes -  
XX  
PS Claim 14; Page 318; 678pp; English.  
XX  
CC The invention relates to a novel DNA array giving a representation of a

CC number of *Methylococcus capsulatus* genes. The method of the invention is  
CC useful for determination of the differential expression of the genes of  
CC *M. capsulatus*, and for studying gene expression on a genomic scale and in  
CC gene expression assays of *M. capsulatus* genes. The sequences shown in  
CC ABQ9016-ABQ91855 represent *M. capsulatus* genes for use in arrays of the  
CC invention.  
XX  
SQ Sequence 225 BP; 56 A; 42 C; 72 G; 55 T; 0 other;  
Query Match 15.0%; Score 19; DB 24; Length 225;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 TACCTGTGCATCTGCAGG 25  
Db 31 TACCTGTGCATCTGCAGG 49  
RESULT 37  
ID AAL36982  
XX AAL36982 standard; DNA; 498 BP.  
AC AAL36982;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3347.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.

08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0250100.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251979.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02559678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-451937/48.  
DR  
XX  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the musculoskeletal system including  
PT musculoskeletal cancers and also for testing and detection e.g.  
PT diagnosis -  
XX  
XX Example 2; SEQ ID NO 3347; 781bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins  
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful  
CC for preventing, treating or ameliorating medical conditions e.g. by  
CC protein or gene therapy. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
CC other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIFO at ftp.wifo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 498 BP; 115 A; 115 C; 109 G; 159 T; 0 other;  
SQ  
Query Match 15.0%; Score 19; DB 22; Length 498;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 AACACGACAGGACTCT 82  
Db 403 AACACGACAGGACTCT 421  
RESULT 38  
ID ABX59970 standard; cDNA; 498 BP.  
XX ABX59970;  
AC  
XX  
XX  
XX 26-FEB-2003 (first entry)  
DE  
XX  
XX cDNA encoding novel human musculoskeletal system antigen #2314.  
XX  
XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
XX re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
XX cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
XX post-operative tissue repair; limb regeneration; neuronal growth;  
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
XX AIDS-related complex; chondrocyte growth; bone regeneration;  
XX periodontal regeneration; tissue transport; bone graft; skin aging;  
XX keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
XX cell growth; organ transplant; cell differentiation; body height;  
XX weight; hair colour; eye colour; skin; percentage of adipose tissue;  
XX pigmentation; cosmetic surgery; metabolism; biorhythm; circadian rhythm;  
XX depression; tendency for violence; pain; reproductive capability;  
XX hormone level; endocrine level; appetite; libido; memory; stress;  
XX storage capability; fat content; lipid content; protein content;  
XX carbohydrate content; vitamin content; cofactor content;  
XX nutritional component.  
OS Homo sapiens.  
XX  
XX US2002147140-A1.  
XX  
XX 10-OCT-2002.  
XX  
XX 17-JAN-2001; 2001US-0764877.  
XX  
XX 31-JAN-2000; 2000US-179065P.  
XX 04-FEB-2000; 2000US-180628P.  
XX 28-JUN-2000; 2000US-214886P.  
XX 07-JUL-2000; 2000US-216647P.  
XX 07-JUL-2000; 2000US-216880P.  
XX 11-JUL-2000; 2000US-217487P.  
XX 14-JUL-2000; 2000US-218290P.  
XX 26-JUL-2000; 2000US-220963P.  
XX 26-JUL-2000; 2000US-220964P.  
XX 14-AUG-2000; 2000US-224518P.  
XX 14-AUG-2000; 2000US-224519P.  
XX 14-AUG-2000; 2000US-225267P.  
XX 14-AUG-2000; 2000US-225268P.  
XX 14-AUG-2000; 2000US-225470P.  
XX 14-AUG-2000; 2000US-225471P.  
XX 14-AUG-2000; 2000US-225757P.  
XX 14-AUG-2000; 2000US-225758P.  
XX 22-AUG-2000; 2000US-226868P.  
XX 30-AUG-2000; 2000US-228924P.  
XX 01-SEP-2000; 2000US-229287P.  
XX 01-SEP-2000; 2000US-229287P.  
XX 01-SEP-2000; 2000US-229344P.  
XX 01-SEP-2000; 2000US-229345P.  
XX 01-SEP-2000; 2000US-229345P.  
XX 05-SEP-2000; 2000US-229509P.  
XX 05-SEP-2000; 2000US-229513P.  
XX 08-SEP-2000; 2000US-231413P.  
XX 21-SEP-2000; 2000US-234223P.  
XX 21-SEP-2000; 2000US-234274P.  
XX 25-SEP-2000; 2000US-234997P.  
XX 27-SEP-2000; 2000US-235834P.  
XX 29-SEP-2000; 2000US-236327P.  
XX 29-SEP-2000; 2000US-236367P.

PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-239355P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBEN/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 DR WPI; 2003-128199/12.  
 XX  
 PT Isolated nucleic acid molecules encoding musculoskeletal system  
 associated polypeptides, useful for detecting disorders, e.g., cancer -  
 XX  
 PS Disclosure; SEQ ID NO 3347; 321bp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule comprising a  
 CC sequence encoding musculoskeletal system associated polypeptides useful  
 CC for detecting disorders, e.g., cancer or cancer metastases, in animals  
 CC or humans. The nucleic acid; stimulates re-vascularisation of ischaemic  
 CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
 CC and other cardiovascular conditions; treats wounds due to injuries;  
 CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
 CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
 CC neuronal damage occurring in certain disorders or neurodegenerative  
 CC conditions, such as, Alzheimer's disease, Parkinson's disease, and  
 CC AIDS-related complex; stimulates chondrocyte growth, thus they can be  
 CC used to enhance bone and periodontal regeneration and aid in tissue  
 CC transports or bone grafts; prevents skin aging due to sunburn by  
 CC stimulating keratinocyte growth; prevents hair loss, since FGF family  
 CC members activate hair-forming cells and promotes melanocyte growth;  
 CC stimulates growth and differentiation of hematopoietic cells and bone  
 CC marrow cells when used in combination with other cytokines; maintains  
 CC organs before transplantation or for supporting cell culture of primary  
 CC tissues; induces tissue of mesodermal origin to differentiate in early  
 CC embryos; increases or decreases the differentiation or proliferation of  
 CC embryonic stem cells; besides, haematopoietic lineage; modulates  
 CC mammalian characteristics, such as, body height, weight, hair colour, eye  
 CC colour, skin, percentage of adipose tissue, pigmentation, size, and shape  
 CC (e.g., cosmetic surgery); modulates mammalian metabolism; changes  
 CC mammal's metal state or physical state by influencing biorhythms,  
 CC cardiac rhythms, depression, tendency for violence, tolerance for pain,  
 CC reproductive capabilities, hormonal or endocrine levels, appetite,  
 CC libido, memory, or stress; increases or decreases storage capabilities,  
 CC fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors  
 CC or other nutritional components. This sequence encodes a novel human  
 CC musculoskeletal system antigen.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from the US patent office at  
 CC ftp.segdata.uspto.gov/sequence.html?docID=20020147140.  
 XX  
 SQ Sequence 498 BP; 115 A; 115 C; 109 G; 159 T; 0 other;  
 Query Match 15.0%; Score 19; DB 25; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AACACAGAGAGACTGT 82  
 Db 403 AACACAGAGAGACTGT 421  
 RESULT 39  
 ID AA87337 standard; DNA; 18 BP.  
 XX  
 AC AA87337;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX  
 DE Human sensitive to apoptosis gene (SAG) primer hSAG.M1.  
 XX SAG gene; sensitive to apoptosis; human; cancer; tumour;  
 KW neurodegenerative disease; muscular dystrophy; wound healing;  
 KW vulnerability; therapy; PCR; primer; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FN WO9932514-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PF 15-DEC-1998; 98WO-US26705.  
 XX  
 PF 11-SEP-1998; 98US-0099840.  
 PR 19-DEC-1997; 97US-0068179.  
 XX  
 PA (WARN ) WARNER LAMBERT CO.  
 XX  
 PI Sun Y;  
 XX  
 DR WPI; 1999-430152/36.  
 XX  
 PT SAG: Sensitive to Apoptosis Gene and related proteins, useful for  
 PT promoting cell growth and protecting cells against apoptosis  
 XX  
 PS Example 13; Page 26; 84bp; English.  
 XX  
 CC Primer hSAG.M1 corresponds to nucleotides 151-171 of the human  
 CC sensitive to apoptosis gene (SAG) cDNA clone provided in AA87314.  
 CC It was used with primer SAGT.02-1 (see AA87338) in the RT-PCR  
 CC amplification of RNA isolated from 20 human tumour lines and  
 CC transformed lines originating from lung, brain, kidney, prostate,  
 CC testis, nasopharynx, bone, cervix and foreskin. 2 SAG deletion  
 CC mutants (see AA87315-16) were detected in cancer cell lines  
 CC originating from colon and testis, suggesting a possible role for  
 CC SAG in human carcinogenesis. SAG (see also AA106492) is a  
 CC redox-sensitive, haem-binding protein that promotes cell growth,  
 CC protects cells from apoptosis, and scavenges oxygen radicals. It  
 CC can be used to reverse a tumour phenotype.  
 XX  
 SQ Sequence 18 BP; 3 A; 6 C; 6 G; 3 T; 0 other;  
 Query Match 14.2%; Score 18; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 14 GCCATTCGACGGTCCAG 31  
 Db 1 GCCATTCGACGGTCCAG 18  
 RESULT 40  
 ID AA84630 standard; cDNA; 239 BP.  
 XX  
 AC AA84630;  
 XX  
 DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #20434.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX Food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS  
PN MO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
PA (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT,  
PI WPI; 2001-639362/73.  
DR P-PSDB; ABG20443.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID NO 20434; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensic, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_poc\_sequences.  
XX  
XX Sequence 239 BP; 47 A; 87 C; 58 G; 47 T; 0 other;  
SQ  
XX  
XX Query Match 13.4%; Score 17; DB 23; Length 239;  
XX Best Local Similarity 100.0%; Pred. No. 39;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 CCATCTGCAGGGTCCAG 31  
DB 139 CCATCTGCAGGGTCCAG 123

RESULT 41  
ABX20540/c  
ID ABX20540 standard; cDNA; 348 BP.  
XX  
XX ABX20540;  
AC  
XX 10-FEB-2003 (first entry)  
DT  
XX Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #2597.  
DE

XX Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;  
XX cellular fucosylation; glycoconjugate fucosylation; transplant rejection;  
XX arthritis; asthma; sepsis; reperfusion injury; stroke; infection;  
XX complex carbohydrate; gene replacement therapy; immunosuppressive;  
XX antiinflammatory; antibacterial; antithyroid; cerebroprotective;  
XX antiasthmatic; vasotropic.  
XX  
OS Homo sapiens.  
PN US2002110548-A1.  
XX  
PD 15-AUG-2002.  
XX  
PF 11-JUN-2001; 2001US-0878574.  
XX  
PR 22-NOV-1996; 96US-0753233.  
XX 03-DEC-1997; 97US-0984246.  
PR 09-SEP-1998; 98US-0149674.  
PR 14-JUN-1999; 99US-0333177.  
XX  
PA (GENV) GENETICS INST INC.  
PI  
PI Sullivan F, Kriz R, Kumar R;  
PI WPI; 2003-066673/06.  
DR  
XX  
XX New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)  
PT peptide, for manufacturing complex carbohydrates, or as targets for  
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant  
PT rejection -  
XX  
XX Disclosure; SEQ ID NO 2599; 6pp; English.  
PS  
XX  
XX The invention relates to a composition comprising a human GDP-mannose  
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying  
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation  
CC in a mammalian subject and for treating or ameliorating diseases affected  
CC by the level of cellular fucosylation or diseases affected by the  
CC fucosylation of glycoconjugates. These diseases include arthritis,  
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or  
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also  
CC useful for manufacturing complex carbohydrates and as targets for  
CC screening small molecule antagonists of the activity of the enzyme. The  
CC polynucleotide is useful in developing an assay for defects in the  
CC enzyme, as well as in gene replacement therapy. Sequences  
CC ABX17942-ABX17944 and ABX17947-ABX3716 represent DNA molecules encoding  
CC human GM4,6D peptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html.  
XX  
XX Sequence 348 BP; 114 A; 56 C; 75 G; 103 T; 0 other;  
SQ  
XX  
XX Query Match 13.4%; Score 17; DB 25; Length 348;  
XX Best Local Similarity 100.0%; Pred. No. 39;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 TGTATCAATTCCTTCCA 117  
DB 205 TGTATCAATTCCTTCCA 189

RESULT 42  
ABA46329  
ID ABA46329 standard; DNA; 355 BP.  
XX  
XX ABA46329;  
AC  
XX 01-FEB-2002 (first entry)  
DT  
XX Human breast cell single exon nucleic acid probe #5024.  
DE

```

KM Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157271-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JUN-2001; 2001WO-US00662.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-496933/54.
XX
XX PT New spatially-addressable set of single exon nucleic acid probes,
XX PT useful for measuring gene expression in sample derived from human
XX PT breast, comprises number of single exon nucleic acid probes
XX
XX PS Claim 1; SEQ ID NO 5024; 327bp + sequence listing; English.
XX
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting
XX CC the probes with a collection of detectably labelled nucleic acids
XX CC derived from mRNA of human breast, and then measuring the label
XX CC bound to each probe of the microarray. The probes are useful for
XX CC verifying the expression of regions of genomic DNA predicted to
XX CC encode proteins. They are useful for gene discovery, and for
XX CC determining predisposition and/or prognosing breast disease. Gene
XX CC expression analysis is useful for assessing the toxicity of chemical
XX CC agents on cells. The microarray of this invention presents a far greater
XX CC diversity of probes for measuring gene expression, with far less bias
XX CC than expressed sequence tag microarrays. The method is suitable for
XX CC rapid production of functional information from genomic sequence. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;
XX
XX Query Match 13.4%; Score 17; DB 22; Length 355;
XX Best Local Similarity 100.0%; Pred. No. 39;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 25 GGTCGAGGTGATGATG 41
XX ||||||||||||||||
XX Db 281 GGTCGAGGTGATGATG 297
XX
XX RESULT 43
XX ID ABA51431 standard; DNA; 355 BP.
XX
XX AC ABA51431;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human breast cell single exon nucleic acid probe #10126.
XX
XX KM Human; microarray; single exon probe; gene expression; breast;
XX KM disease; cancer; ss.

```

```

XX
XX OS Homo sapiens.
XX
XX PN WO200157271-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JUN-2001; 2001WO-US00662.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-496933/54.
XX
XX PT New spatially-addressable set of single exon nucleic acid probes,
XX PT useful for measuring gene expression in sample derived from human
XX PT breast, comprises number of single exon nucleic acid probes
XX
XX PS Claim 4; SEQ ID NO 10126; 327bp + sequence listing; English.
XX
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting
XX CC the probes with a collection of detectably labelled nucleic acids
XX CC derived from mRNA of human breast, and then measuring the label
XX CC bound to each probe of the microarray. The probes are useful for
XX CC verifying the expression of regions of genomic DNA predicted to
XX CC encode proteins. They are useful for gene discovery, and for
XX CC determining predisposition and/or prognosing breast disease. Gene
XX CC expression analysis is useful for assessing the toxicity of chemical
XX CC agents on cells. The microarray of this invention presents a far greater
XX CC diversity of probes for measuring gene expression, with far less bias
XX CC than expressed sequence tag microarrays. The method is suitable for
XX CC rapid production of functional information from genomic sequence. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;
XX
XX Query Match 13.4%; Score 17; DB 22; Length 355;
XX Best Local Similarity 100.0%; Pred. No. 39;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 25 GGTCGAGGTGATGATG 41
XX ||||||||||||||||
XX Db 281 GGTCGAGGTGATGATG 297
XX
XX RESULT 44
XX ID ABA56889 standard; DNA; 355 BP.
XX
XX AC ABA56889;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #5194.
XX
XX KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX OS Homo sapiens.
XX

```



PN WO200157277-A2.  
XX  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX DR  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX  
XX PS Claim 1; SEQ ID NO 5194; 639bp + sequence listing; English.  
XX  
XX CC The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX fetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.  
XX CC Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
XX  
XX  
XX Query Match 13.4%; Score 17; DB 22; Length 355;  
XX Best Local Similarity 100.0%; Pred. No. 39;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 25 GGTCGAGTGATGATG 41  
XX |||||  
XX DB 281 GGTCGAGTGATGATG 297  
XX  
XX  
XX RESULT 45  
XX ABA69454  
XX ID ABA69454 standard; DNA; 355 BP.  
XX  
XX AC ABA69454;  
XX  
XX DT 01-FEB-2002 (first entry)  
XX  
XX DE Human foetal liver single exon nucleic acid probe #17759.  
XX  
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157277-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00669.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX

XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX DR  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX  
XX PS Claim 4; SEQ ID NO 17759; 639bp + sequence listing; English.  
XX  
XX CC The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX fetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.  
XX CC Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
XX  
XX  
XX Query Match 13.4%; Score 17; DB 22; Length 355;  
XX Best Local Similarity 100.0%; Pred. No. 39;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 25 GGTCGAGTGATGATG 41  
XX |||||  
XX DB 281 GGTCGAGTGATGATG 297  
XX  
XX  
XX RESULT 46  
XX ABA26508  
XX ID ABA26508 standard; DNA; 355 BP.  
XX  
XX AC ABA26508;  
XX  
XX DT 23-JAN-2002 (first entry)  
XX  
XX DE Probe #4974 for gene expression analysis in human heart cell sample.  
XX  
XX KW Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157274-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00666.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX DR  
XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX PS Claim 1; SEQ ID NO 4974; 530bp; English.  
XX

XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
  
Query Match 13.4%; Score 17; DB 22; Length 355;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 25 GGTCCAGGTGATGATG 41  
DB 281 GGTCCAGGTGATGATG 297  
  
RESULT 47  
ABA36383  
ID ABA36383 standard; DNA; 355 BP.  
XX ABA36383;  
XX  
XX 23-JAN-2002 (first entry)  
XX  
DE Probe #14849 for gene expression analysis in human heart cell sample.  
XX  
XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00666.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX Claim 4; SEQ ID No 14849; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging,  
XX monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
  
Query Match 13.4%; Score 17; DB 22; Length 355;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 25 GGTCCAGGTGATGATG 41  
DB 281 GGTCCAGGTGATGATG 297  
  
RESULT 48  
AAK04990  
ID AAK04990 standard; DNA; 355 BP.  
XX AAK04990;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 4981.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
XX Example 4; SEQ ID NO: 4981; 650pp + Sequence listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention.  
XX  
SQ Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;  
  
Query Match 13.4%; Score 17; DB 22; Length 355;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 25 GGTCCAGGTGATGATG 41

Db 281 GGTCGAGTGATGATG 297

## RESULT 49

AAK17713

AAK17713 standard; DNA; 355 BP.

AAK17713;

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 17704.

Human; brain expressed exon; gene expression analysis; probe;

microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

epilepsy; cancer; ss.

Homo sapiens.

MO200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00667.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human

Example 4; SEQ ID NO: 17704; 650bp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid

probes which are derived from genomic sequences expressed in the human

brain. They can be used to measure gene expression in brain cell samples,

which may enable the diagnosis and improved treatment of nervous system

diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

epilepsy and cancers. The present sequence is one of the probes of the

invention.

Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;

Query Match 13.4%; Score 17; DB 22; Length 355;

Best Local Similarity 100.0%; Pred. No. 39; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 GGTCGAGTGATGATG 41

Db 281 GGTCGAGTGATGATG 297

## RESULT 50

AAK30523

AAK30523 standard; DNA; 355 BP.

AAK30523;

06-NOV-2001 (first entry)

Human bone marrow expressed single exon probe SEQ ID NO: 5080.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human bone marrow -

Example 4; SEQ ID NO: 5080; 658bp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid

probes which are derived from genomic sequences expressed in the human

bone marrow. They can be used to measure gene expression in bone marrow

samples, which may enable the improved diagnosis and treatment of cancers

such as lymphoma, leukaemia and myeloma. The present sequence is one of

the probes of the invention.

Sequence 355 BP; 122 A; 47 C; 109 G; 77 T; 0 other;

Query Match 13.4%; Score 17; DB 22; Length 355;

Best Local Similarity 100.0%; Pred. No. 39; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 GGTCGAGTGATGATG 41

Db 281 GGTCGAGTGATGATG 297

Search completed: November 7, 2003, 11:58:13

UOD time : 131.53 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 05:52:38 ; Search time 34.4064 Seconds  
(without alignments)  
1629.223 Million cell updates/sec

Title: US-09-509-779-1\_COPY\_154\_280

Perfect score: 127  
Sequence: 1 GTGCATACCTGTCATCT.....ATTCTTCACACATGCTGC 127

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents\_NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	13.4	363	4	US-09-328-352-1220
2	17	13.4	942	4	US-09-252-991A-11847
3	17	13.4	978	4	US-09-252-991A-11595
4	17	13.4	1605	4	US-09-252-991A-11780
5	17	13.4	3494	3	US-09-139-802-200
6	17	13.4	3494	4	US-09-659-786-200
7	17	13.4	8654	4	US-08-961-527-98
8	16	12.6	458	4	US-09-149-476-239
9	16	12.6	475	4	US-09-149-476-76
10	16	12.6	1382	2	US-08-596-387B-123
11	16	12.6	1382	4	US-09-067-615-123
12	16	12.6	1382	5	PCT-US95-09816A-123
13	16	12.6	1385	2	US-08-596-387B-121
14	16	12.6	1385	4	US-09-067-615-121
15	16	12.6	1385	5	PCT-US95-09816A-121
16	16	12.6	1508	2	US-08-596-387B-122
17	16	12.6	1508	3	US-08-860-190A-24
18	16	12.6	1508	4	US-09-067-615-122
19	16	12.6	1508	5	PCT-US95-09816A-122
20	16	12.6	3208	4	US-09-780-016-27
21	16	12.6	4713	4	US-09-194-285-7
22	16	12.6	5769	1	US-08-652-971-1
23	16	12.6	5769	2	US-08-991-258A-1
24	16	12.6	5769	3	US-08-769-399-1
25	16	12.6	5769	4	US-08-991-953A-1
26	15	11.8	196	3	US-08-663-082-1
27	15	11.8	219	4	US-09-328-352-2677

C	28	15	11.8	266	4	US-09-313-294A-3641	Sequence 3641, Ap
	29	15	11.8	266	4	US-09-702-705-1559	Sequence 1559, Ap
	30	15	11.8	266	4	US-09-736-457-1559	Sequence 1559, Ap
C	31	15	11.8	275	4	US-09-313-294A-2588	Sequence 2588, Ap
	32	15	11.8	318	4	US-09-016-434-1031	Sequence 1031, Ap
	33	15	11.8	474	1	US-08-412-614-93	Sequence 93, Appl
	34	15	11.8	474	2	US-08-635-761-93	Sequence 93, Appl
	35	15	11.8	474	3	US-09-312-520-93	Sequence 94, Appl
	36	15	11.8	476	1	US-08-412-614-94	Sequence 94, Appl
	37	15	11.8	476	2	US-08-635-761-94	Sequence 94, Appl
	38	15	11.8	476	3	US-09-312-520-94	Sequence 94, Appl
	39	15	11.8	819	1	US-08-309-182B-2	Sequence 2, Appl
	40	15	11.8	1426	3	US-09-121-425-3	Sequence 3, Appl
C	41	15	11.8	1426	4	US-09-634-493A-3	Sequence 3, Appl
C	42	15	11.8	2238	1	US-08-389-668A-1	Sequence 1, Appl
C	43	15	11.8	2238	1	US-08-732-506-1	Sequence 1, Appl
C	44	15	11.8	2238	5	PCT-US95-05768-1	Sequence 1, Appl
C	45	15	11.8	2508	1	US-07-959-943-10	Sequence 10, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-328-352-1220
; Sequence 1220, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIORITY FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1220
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1220

Query Match          13.4%; Score 17; DB 4; Length 363;
Best local similarity 100.0%; Pred. No. 5.8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 TGTGTTGCTGTCGGG 96
      |||||
Db      213 TGTGTTGCTGTCGGG 229

RESULT 2
US-09-252-991A-11847
; Sequence 11847, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,768
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11847
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11847

Query Match          13.4%; Score 17; DB 4; Length 942;
```

Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 GGTGATGATGCTGCC 47  
Db 489 GGTGATGATGCTGCC 505

## RESULT 3

US-09-252-991A-11595/c  
; Sequence 11595, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11595  
; LENGTH: 978  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11595

Query Match 13.4%; Score 17; DB 4; Length 978;  
Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 GGTGATGATGCTGCC 47  
Db 424 GGTGATGATGCTGCC 408

## RESULT 4

US-09-252-991A-11780  
; Sequence 11780, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11780  
; LENGTH: 1605  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11780

Query Match 13.4%; Score 17; DB 4; Length 1605;  
Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 GGTGATGATGCTGCC 47  
Db 409 GGTGATGATGCTGCC 425

## RESULT 5

US-09-139-802-200/c  
; Sequence 200, Application US/09139802

Patent No. 6180084  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
; FILE REFERENCE: P-LJ 3203  
; CURRENT APPLICATION NUMBER: US/09/139,802  
; PRIOR FILING DATE: 1998-08-25  
; EARLIER APPLICATION NUMBER: 08/926,914  
; EARLIER FILING DATE: 1997-09-10  
; EARLIER APPLICATION NUMBER: 08/710,067  
; NUMBER OF SEQ ID NOS: 226  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 200  
; LENGTH: 3494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (121)..(3024)  
US-09-139-802-200

Query Match 13.4%; Score 17; DB 3; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 6;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 CCATCTGAGGGTCCAG 31  
Db 1762 CCATCTGAGGGTCCAG 1746

## RESULT 6

US-09-659-786-200/c  
; Sequence 200, Application US/09659786  
; Patent No. 6491894  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
; FILE REFERENCE: P-LJ 3203  
; CURRENT APPLICATION NUMBER: US/09/659,786  
; PRIOR FILING DATE: 2000-09-11  
; PRIOR APPLICATION NUMBER: 08/926,914  
; PRIOR FILING DATE: 1997-09-10  
; PRIOR APPLICATION NUMBER: 08/710,067  
; PRIOR FILING DATE: 1996-09-10  
; NUMBER OF SEQ ID NOS: 226  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 200  
; LENGTH: 3494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (121)..(3024)  
US-09-659-786-200

Query Match 13.4%; Score 17; DB 4; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 6;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 CCATCTGAGGGTCCAG 31  
Db 1762 CCATCTGAGGGTCCAG 1746

## RESULT 7

US-08-961-527-98/c

Sequence 98, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,33  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8654 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-98

Query Match 13.4%; Score 17; DB 4; Length 8654;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GACTGTGTGTGCTG 93  
DB 1105 GACTGTGTGTGCTG 1089

RESULT 8  
US-09-149-476-239/C  
Sequence 239, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
EARLIER FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,500  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,587  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,492  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,598  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,613  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,582  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,674  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,669  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 12.6%; Score 16; DB 4; Length 458;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 TCTGAGGTCACGCT 33  
Db 238 TCTGAGGTCACGCT 223

RESULT 9  
US-09-149-476-76/c  
Sequence 76, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597



[illegible]

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 12.6%; Score 16; DB 4; Length 475;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TCTGCGGCTCCAGGT 33  
DB 226 TCTGCGGCTCCAGGT 211

RESULT 10  
US-08-596-387B-123  
Sequence 123, Application US/08596387B  
Patent No. 5869270  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Jiao, Jin-An  
APPLICANT: Burkhardt, Martin  
APPLICANT: Wong, Hing  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,387B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09816  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.

REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELEPHONE/COMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1382 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1382  
US-08-596-387B-123

Query Match 12.6%; Score 16; DB 2; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATCTCTCCACAA 120  
DB 1243 ACCATCTCTCCACAA 1258

RESULT 11  
US-09-067-615-123  
Sequence 123, Application US/09067615  
Patent No. 6309645  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Jiao, Jin-An  
APPLICANT: Burkhardt, Martin  
APPLICANT: Wong, Hing  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,615  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/596,387  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1382 base pairs  
TYPE: nucleic acid

STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1382  
US-09-067-615-123

Query Match 12.6%; Score 16; DB 4; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCTTCACAA 120  
DB 1243 ACCATTCTTCACAA 1258

RESULT 12  
PCT-US95-09816A-123  
Sequence 123, Application PC/TUS9509816A  
GENERAL INFORMATION:  
APPLICANT: Wong, Hing C.  
APPLICANT: Rhode, Peter R.  
APPLICANT: Widanz, Jon A.  
APPLICANT: Grammer, Susan  
APPLICANT: Edwards, Ana C.  
APPLICANT: Chavallaz, Pierre-Andre  
APPLICANT: Jiao, Jin-An  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 123  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09816A  
FILING DATE: 31-JUL-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1382 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1382  
PCT-US95-09816A-123  
Query Match 12.6%; Score 16; DB 5; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 20;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 105 ACCATTCTTCACAA 120  
DB 1243 ACCATTCTTCACAA 1258

RESULT 13  
US-08-596-387B-121  
Sequence 121, Application US/08596387B  
Patent No. 5869270  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Burkhardt, Martin  
APPLICANT: Wong, Hing  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,387B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09816  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1382  
US-08-596-387B-121  
Query Match 12.6%; Score 16; DB 2; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 105 ACCATTCTTCACAA 120  
DB 1243 ACCATTCTTCACAA 1258  
RESULT 14  
US-09-067-615-121

; Sequence 121, Application US/09067615  
; Patent No. 6309645  
; GENERAL INFORMATION:  
; APPLICANT: Rhode, Peter R.  
; APPLICANT: Jiao, Jin-An  
; APPLICANT: Burkhardt, Martin  
; APPLICANT: Wong, Hing  
; TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dade International, Inc.  
; STREET: 1717 Deerfield Road  
; CITY: Deerfield  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60015  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,615  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/596,387  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/382,454  
; FILING DATE: 01-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,302  
; FILING DATE: 29-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pearson, Louise S.  
; REGISTRATION NUMBER: 32,369  
; REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (708) 267-5376  
; TELEFAX: (708) 267-5376  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1385 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; FEATURE:  
; MOLECULE TYPE: DNA (genomic)  
; NAME/KEY: CDS  
; LOCATION: 6..1382  
; US-09-067-615-121

Query Match 12.6%; Score 16; DB 4; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 ACCATTCCTTCACAA 120  
|||  
Db 1243 ACCATTCCTTCACAA 1258

RESULT 15  
PCT-US95-09816A-121  
; Sequence 121, Application PC/TUS9509816A  
; GENERAL INFORMATION:  
; APPLICANT: Mong, Hing C.  
; APPLICANT: Rhode, Peter R.  
; APPLICANT: Widanz, Jon A.  
; APPLICANT: Grammer, Susan  
; APPLICANT: Edwards, Ana C.  
; APPLICANT: Chavalliaz, Pierre-Andre  
; APPLICANT: Jiao, Jin-An

; TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
; NUMBER OF SEQUENCES: 123  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dade International, Inc.  
; STREET: 1717 Deerfield Road  
; CITY: Deerfield  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60015  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/09816A  
; FILING DATE: 31-JUL-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/382,454  
; FILING DATE: 01-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,302  
; FILING DATE: 29-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pearson, Louise S.  
; REGISTRATION NUMBER: 32,369  
; REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (708) 267-5376  
; TELEFAX: (708) 267-5376  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1385 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 6..1382  
; PCT-US95-09816A-121

Query Match 12.6%; Score 16; DB 5; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 ACCATTCCTTCACAA 120  
|||  
Db 1243 ACCATTCCTTCACAA 1258

RESULT 16  
US-08-596-387B-122  
; Sequence 122, Application US/08596387B  
; Patent No. 5869270  
; GENERAL INFORMATION:  
; APPLICANT: Rhode, Peter R.  
; APPLICANT: Jiao, Jin-An  
; APPLICANT: Burkhardt, Martin  
; APPLICANT: Wong, Hing  
; TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dade International, Inc.  
; STREET: 1717 Deerfield Road  
; CITY: Deerfield  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60015  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,387B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09816  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1505  
US-08-596-387B-122

Query Match 12.6%; Score 16; DB 2; Length 1508;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCTTCCACAA 120  
DB 1243 ACCATTCTTCCACAA 1258

RESULT 17  
US-08-960-190A-24  
Sequence 24, Application US/08960190A  
Patent No. 6232445  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Acevedo, Jorge  
APPLICANT: Burkhardt, Martin  
APPLICANT: Jiao, Jin-an  
APPLICANT: Wong, Hing C.  
TITLE OF INVENTION: SOLUBLE MHC COMPLEXES AND  
METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: usa  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,190A  
FILING DATE: 29-OCT-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cortles, Peter F.  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 48002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 6..1505  
OTHER INFORMATION:  
US-08-960-190A-24

Query Match 12.6%; Score 16; DB 3; Length 1508;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCTTCCACAA 120  
DB 1243 ACCATTCTTCCACAA 1258

RESULT 18  
US-09-067-615-122  
Sequence 122, Application US/09067615  
Patent No. 6309645  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Burkhardt, Martin  
APPLICANT: Jiao, Jin-an  
APPLICANT: Wong, Hing  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,615  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/596,387  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300

TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6..1505  
US-09-067-615-122

Query Match 12.6%; Score 16; DB 4; Length 1508;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCCTTCACAA 120  
DB 1243 ACCATTCCTTCACAA 1258

RESULT 19  
PCT-US95-09816A-122  
Sequence 122, Application PC/TUS9509816A  
GENERAL INFORMATION:  
APPLICANT: Wong, Hing C.  
APPLICANT: Rhoads, Peter R.  
APPLICANT: Widanz, Jon A.  
APPLICANT: Grammer, Susan  
APPLICANT: Edwards, Ana C.  
APPLICANT: Chavallaz, Pierre-Andre  
APPLICANT: Jiao, Jin-An  
TITLE OF INVENTION: MHC COMPLEXES AND USES THEREOF  
NUMBER OF SEQUENCES: 123  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade International, Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09816A  
FILING DATE: 31-JUL-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/382,454  
FILING DATE: 01-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,302  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Pearson, Louise S.  
REGISTRATION NUMBER: 32,369  
REFERENCE/DOCKET NUMBER: STR-4665-CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 267-5300  
TELEFAX: (708) 267-5376  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS

LOCATION: 6..1505  
PCT-US95-09816A-122

Query Match 12.6%; Score 16; DB 5; Length 1508;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCCTTCACAA 120  
DB 1243 ACCATTCCTTCACAA 1258

RESULT 20  
US-09-780-016-27  
Sequence 27, Application US/09780016  
Patent No. 6509456  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Scoville, John  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
TITLE OF INVENTION: No. 6509456el Human Proteases and  
TITLE OF INVENTION: Polynucleotides Encoding the Same  
FILE REFERENCE: LEX-0132-USA  
CURRENT APPLICATION NUMBER: US/09/780,016  
CURRENT FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: US 60/181,294  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 3208  
TYPE: DNA  
ORGANISM: homo sapiens  
US-09-780-016-27

Query Match 12.6%; Score 16; DB 4; Length 3208;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 CTGTGCATCTGCAGG 25  
DB 2820 CTGTGCATCTGCAGG 2835

RESULT 21  
US-09-194-285-7  
Sequence 7, Application US/09194285  
Patent No. 6355479  
GENERAL INFORMATION:  
APPLICANT: Webb, Susan R.  
APPLICANT: Winovist, Ola  
APPLICANT: Karlsson, Lars  
APPLICANT: Jackson, Michael R.  
TITLE OF INVENTION: MHC Class II Antigen Presenting Systems  
TITLE OF INVENTION: And Methods for Activating CD4+ T Cells  
FILE REFERENCE: TSRI 536.1  
CURRENT APPLICATION NUMBER: US/09/194,285  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: PCT/US97/08697  
PRIOR FILING DATE: 1997-05-22  
PRIOR APPLICATION NUMBER: US 60/018,175  
PRIOR FILING DATE: 1996-05-23  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 4713  
TYPE: DNA  
ORGANISM: Mus musculus

US-09-194-285-7

Query Match 12.6%; Score 16; DB 4; Length 4713;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ACCATTCCTCCACAA 120

DB 964 ACCATTCCTCCACAA 979

RESULT 22

US-08-652-971-1/c

; Sequence 1, Application us/08652971

; Patent No. 5914507

; GENERAL INFORMATION:

; APPLICANT: Cheng, Jill

; APPLICANT: Lasky, Laurence A.

; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE

; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd.

; CITY: South San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/652,971

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Dregger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1033

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 225-3216

; TELEFAX: (415) 952-9881

; TELEX: 910 371-7168

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5769 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 379..4686

; US-08-652-971-1

QY 84 TTGTGCTCGGGAGA 99

DB 890 TTGTGCTCGGGAGA 875

RESULT 23

US-08-991-258A-1/c

; Sequence 1, Application US/08991258A

; Patent No. 5928887

; GENERAL INFORMATION:

; APPLICANT: Cheng, Jill

; APPLICANT: Lasky, Laurence A.

; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE

; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON &amp; HERBERT, LLP

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US 08/652,971

; FILING DATE: 17-DEC-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 24-MAY-1996

; FILING DATE: 24-MAY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Dregger, Walter H.

; REGISTRATION NUMBER: 24,190

; REFERENCE/DOCKET NUMBER: A-63478-3/WH/MTK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5769 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 379..4686

; US-08-991-258A-1

QY 84 TTGTGCTCGGGAGA 99

DB 890 TTGTGCTCGGGAGA 875

RESULT 24

US-08-769-399-1/c

; Sequence 1, Application US/08769399

; Patent No. 5976852

; GENERAL INFORMATION:

; APPLICANT: Cheng, Jill

; APPLICANT: Lasky, Laurence A.

; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE

; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd.

; CITY: South San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/769,399  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dreger, Ginger R.  
;; REGISTRATION NUMBER: 33,055  
;; REFERENCE/DOCKET NUMBER: PI033  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 225-3216  
;; TELEFAX: (415) 952-9881  
;; TELEEX: 910 371-7168  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5769 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 379..4686  
;;  
US-08-769-399-1  
  
Query Match 12.6%; Score 16; DB 2; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 84 TTGTGCTGTGGGAGAGA 99  
|||  
DB 890 TTGTGCTGTGGGAGAGA 875  
  
RESULT 25  
US-08-991-953A-1/c  
; Sequence 1, Application US/08991953A  
; Patent No. 6083748  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT, LLP  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/991,953A  
; FILING DATE: 16-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/652,971  
; FILING DATE: 24-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63478-3/WHO/MTK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5769 base pairs

;;  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 379..4686  
;;  
US-08-991-953A-1  
  
Query Match 12.6%; Score 16; DB 3; Length 5769;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 84 TTGTGCTGTGGGAGAGA 99  
|||  
DB 890 TTGTGCTGTGGGAGAGA 875  
  
RESULT 26  
US-08-663-082-1  
; Sequence 1, Application US/08663082  
; Patent No. 6043411  
; GENERAL INFORMATION:  
; APPLICANT: NISHIZAWA, Osamu  
; APPLICANT: TOGURI, Toshihiro  
; TITLE OF INVENTION: GENE FOR FATTY ACID DESATURASE, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID GENE, PLANT TRANSFORMED WITH SAID GENE,  
; TITLE OF INVENTION: AND PROCESS FOR CREATING SAID PLANT  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,082  
; FILING DATE: 25-JUN-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP94/02288  
; FILING DATE: 28-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 93/352858  
; FILING DATE: 28-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 81356/107  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: Anabaena variabilis  
; STRAIN: IAM M-3  
;;  
US-08-663-082-1  
  
Query Match 11.8%; Score 15; DB 3; Length 196;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 84 TTGTGGCTGGGAG 98  
|||||  
Db 41 TTGTGGCTGGGAG 55

RESULT 27  
US-09-328-352-2677  
; Sequence 2677, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 2677  
; LENGTH: 219  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-2677

Query Match 11.8%; Score 15; DB 4; Length 219;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 60 CTGAAACAGCAG 74  
|||||  
Db 101 CTGAAACAGCAG 115

RESULT 28  
US-09-313-294A-3641/C  
; Sequence 3641, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 3641  
; LENGTH: 266  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700612036H1  
; NAME/KEY: unsure  
; LOCATION: 128  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-3641

Query Match 11.8%; Score 15; DB 4; Length 266;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 51 GATGTCAGCTGAA 65  
|||||  
Db 154 GATGTCAGCTGAA 140

RESULT 29  
US-09-702-705-1559  
; Sequence 1559, Application US/09702705  
; Patent No. 6504010  
; GENERAL INFORMATION:

; APPLICANT: Wang, Tongrong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darriek  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.478C14  
; CURRENT APPLICATION NUMBER: US/09/702,705  
; CURRENT FILING DATE: 2000-10-30  
; NUMBER OF SEQ ID NOS: 1833  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1559  
; LENGTH: 266  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-702-705-1559

Query Match 11.8%; Score 15; DB 4; Length 266;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 GCTGAAACAGCAA 73  
|||||  
Db 196 GCTGAAACAGCAA 210

RESULT 30  
US-09-736-457-1559.  
; Sequence 1559, Application US/09736457  
; Patent No. 6509448  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongrong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darriek  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; APPLICANT: Wang, Aijun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.478C15  
; CURRENT APPLICATION NUMBER: US/09/736,457  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 1864  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1559  
; LENGTH: 266  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-736-457-1559

Query Match 11.8%; Score 15; DB 4; Length 266;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 GCTGAAACAGCAA 73  
|||||  
Db 196 GCTGAAACAGCAA 210

RESULT 31  
US-09-313-294A-2588/C  
; Sequence 2588, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:

APPLICANT: Lalgudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO: 2588  
LENGTH: 275  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6476212 700552942H1  
US-09-313-294A-2588

Query Match 11.8%; Score 15; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGGAGCTGTGTT 85  
DB 178 CAAGGAGCTGTGTT 164

RESULT 32  
US-09-016-434-1031  
Sequence 1031 Application US/09016434  
Patent No. 6500938  
GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhammer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: word perfect 6.1 for windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,434  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0002 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1031:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTUT03  
CLONE: 961450

US-09-016-434-1031  
Query Match 11.8%; Score 15; DB 4; Length 318;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 CTGTGCATCTGCAG 24  
DB 205 CTGTGCATCTGCAG 219

RESULT 33  
US-08-412-614-93  
Sequence 93 Application US/08412614  
Patent No. 553638  
GENERAL INFORMATION:  
APPLICANT: Rossau, Rudi  
APPLICANT: Van Heuverswyn, Hugo  
TITLE OF INVENTION: Hybridization Probes Derived from the  
TITLE OF INVENTION: Spacer Region Between the 16S and 23S rRNA Genes for the  
TITLE OF INVENTION: Detection of No. 553638-Viral Microorganisms  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 553638west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402-4131  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5 inch, 1.44 mb capacity  
COMPUTER: IBM PC compatible (Compaq Deskpro 286e)  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect Version #5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/412,614  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: PCT/EP91/00743  
FILING DATE: 18-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB/90901054.3  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A.  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75-USWO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 474 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Haemophilus influenzae  
STRAIN: NCTC 8143  
US-08-412-614-93

Query Match 11.8%; Score 15; DB 1; Length 474;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAACACAG 70  
|||||

Db 258 CAAGCTGAAAAACAAG 272

## RESULT 34

US-08-635-761-93  
Sequence 93, Application US/08635761  
Patent No. 5945282

## GENERAL INFORMATION:

APPLICANT: Rossau, Rudi  
TITLE OF INVENTION: HYBRIDIZATION PROBES DERIVED FROM THE SPACER REGION BE  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 5945282west Center, 90 S. 7th Street  
CITY: Minneapolis  
STATE: MN  
COUNTRY: U.S.A.  
ZIP: 55402

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,761  
FILING DATE: 19-APR-1996  
CLASSIFICATION: 435

## PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: 08/412,614  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75USC1

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 612/332-5300  
TELEFAX: 612/332/9081  
TELEX:

## INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:  
LENGTH: 474 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-635-761-93

## Query Match 11.8%; Score 15; DB 2; Length 474;

Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAAAAACAAG 70  
Db 258 CAAGCTGAAAAACAAG 272

## RESULT 35

US-09-312-520-93  
Sequence 93, Application US/09312520  
Patent No. 6277577

## GENERAL INFORMATION:

APPLICANT: Rossau, Rudi  
TITLE OF INVENTION: HYBRIDIZATION PROBES DERIVED FROM THE SPACER REGION BE  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 6277577west Center, 90 S. 7th Street

CITY: Minneapolis  
STATE: MN  
COUNTRY: U.S.A.  
ZIP: 55402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,520  
FILING DATE: 19-APR-1996  
CLASSIFICATION: 435

## PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: 08/412,614  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75USC1

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 612/332-5300  
TELEFAX: 612/332/9081  
TELEX:

## INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:  
LENGTH: 474 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-09-312-520-93

Query Match 11.8%; Score 15; DB 3; Length 474;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAAAAACAAG 70  
Db 258 CAAGCTGAAAAACAAG 272

## RESULT 36

US-08-412-614-94  
Sequence 94, Application US/08412614  
Patent No. 5536638

## GENERAL INFORMATION:

APPLICANT: Rossau, Rudi  
TITLE OF INVENTION: Hybridization Probes Derived from the  
TITLE OF INVENTION: Spacer Region Between the 16S and 23S rRNA Genes for the  
TITLE OF INVENTION: Detection of No. 5536638-Viral Microorganisms  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 5536638west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402-4131

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette 3.5 inch, 1.44 mb capacity  
COMPUTER: IBM PC compatible (Compaq Deskpro 286e)  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect Version #5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/412,614

FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: PCT/EP91/00743  
FILING DATE: 18-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB/90901054.3  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A.  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75-USWO.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Haemophilus influenzae  
STRAIN: ITM 859  
US-08-412-614-94

Query Match 11.8%; Score 15; DB 1; Length 476;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 CAAGCTGAAACAG 70  
Db 258 CAAGCTGAAACAG 272

RESULT 37  
US-08-635-761-94  
Sequence 94, Application US/08635761  
Patent No. 5945282  
GENERAL INFORMATION:  
APPLICANT: Rossau, Rudi  
TITLE OF INVENTION: HYBRIDIZATION PROBES DERIVED FROM THE SPACER REGION BE  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 5945282west Center, 90 S. 7th Street  
CITY: Minneapolis  
STATE: MN  
COUNTRY: U.S.A.  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,761  
FILING DATE: 19-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: 08/412,614  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A.  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75USC1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/332-5300  
TELEFAX: 612/332/9081  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-635-761-94

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/332-5300  
TELEFAX: 612/332/9081  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-635-761-94

Query Match 11.8%; Score 15; DB 2; Length 476;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 CAAGCTGAAACAG 70  
Db 258 CAAGCTGAAACAG 272

RESULT 38  
US-09-312-520-94  
Sequence 94, Application US/09312520  
Patent No. 627577  
GENERAL INFORMATION:  
APPLICANT: Rossau, Rudi  
TITLE OF INVENTION: HYBRIDIZATION PROBES DERIVED FROM THE SPACER REGION BE  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 627577west Center, 90 S. 7th Street  
CITY: Minneapolis  
STATE: MN  
COUNTRY: U.S.A.  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,520  
FILING DATE: 19-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,394  
FILING DATE: 17-DEC-1992  
APPLICATION NUMBER: 08/412,614  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hillson, Randall A.  
REGISTRATION NUMBER: 31,838  
REFERENCE/DOCKET NUMBER: 8076.75USC1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/332-5300  
TELEFAX: 612/332/9081  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:

ORIGINAL SOURCE:  
US-09-312-520-94

Query Match 11.8%; Score 15; DB 3; Length 476;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAAACACAG 70  
|||||  
DB 258 CAAGCTGAAACACAG 272

RESULT 39

US-08-309-182B-2  
Sequence 2, Application US/08309182B  
Patent No. 5639645

GENERAL INFORMATION:  
APPLICANT: No. 563964510 MURATA  
TITLE OF INVENTION: A RECOMBINANT 9 DESATURASE AND A GENE  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/309,182B  
FILING DATE: September 20, 1994

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 819 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Anabaena variabilis

FEATURE:

NAME/KEY: CDS

LOCATION: 1..816

IDENTIFICATION METHOD: P

US-08-309-182B-2

Query Match 11.8%; Score 15; DB 1; Length 819;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 TTGTGCTGGGAG 98  
|||||  
DB 521 TTGTGCTGGGAG 535

RESULT 40

US-09-121-425-3  
Sequence 3, Application US/09121425  
Patent No. 6153418

GENERAL INFORMATION:  
APPLICANT: Lehmann, Martin  
TITLE OF INVENTION: Consensus Phytases  
FILE REFERENCE: consensus phytases 13239  
CURRENT APPLICATION NUMBER: US/09/121,425  
CURRENT FILING DATE: 1998-07-23  
EARLIER APPLICATION NUMBER: EPO 97112688.3  
EARLIER FILING DATE: 1997-07-24  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3  
LENGTH: 1426  
TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:consensus

US-09-121-425-3

Query Match 11.8%; Score 15; DB 3; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ATGTCAAGCTGAAA 66  
|||||  
DB 1250 ATGTCAAGCTGAAA 1264

RESULT 41

US-09-634-493A-3  
Sequence 3, Application US/09634493A  
Patent No. 6579975

GENERAL INFORMATION:

APPLICANT: Lehmann, Martin

TITLE OF INVENTION: Consensus Phytases

FILE REFERENCE: consensus phytases 13239

CURRENT APPLICATION NUMBER: US/09/634,493A

PRIOR FILING DATE: 2000-08-08

PRIOR APPLICATION NUMBER: US/09/121,425

PRIOR FILING DATE: 1998-07-23

PRIOR APPLICATION NUMBER: EPO 97112688.3

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 1426

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:consensus

US-09-634-493A-3

Query Match 11.8%; Score 15; DB 4; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ATGTCAAGCTGAAA 66  
|||||  
DB 1250 ATGTCAAGCTGAAA 1264

RESULT 42

US-08-389-668A-1/c  
Sequence 1, Application US/08389668A  
Patent No. 5637470

GENERAL INFORMATION:

APPLICANT: Kaczorowski, Gregory J.

APPLICANT: Garcia, Maria L.

APPLICANT: Leonard, Reid J.

APPLICANT: McManus, Owen B.  
APPLICANT: Swanson, Richard J.  
APPLICANT: Folander, Kimberly L.  
TITLE OF INVENTION: NOVEL HETEROMULTIMER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Francis P. Bigley  
STREET: 126 E. Lincoln Avenue, P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/389,668A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/242,811  
FILING DATE: 13-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bigley, Francis P.  
REGISTRATION NUMBER: 36,356  
REFERENCE/DOCKET NUMBER: 19226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-1249  
TELEFAX: (908) 594-4270  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2238 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-08-389-668A-1

Query Match 11.8%; Score 15; DB 1; Length 2238;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AAGCTGAAAACAGC 71  
|||||  
Db 1106 AAGCTGAAAACAGC 1092

RESULT 43  
US-08-732-506-1/c  
Sequence 1, Application US/08732506  
Patent No. 5776734  
GENERAL INFORMATION:  
APPLICANT: Kaczorowski, Gregory J.  
APPLICANT: Garcia, Maria L.  
APPLICANT: Leonard, Reid J.  
APPLICANT: McManus, Owen B.  
APPLICANT: Swanson, Richard J.  
APPLICANT: Folander, Kimberly L.  
TITLE OF INVENTION: NOVEL HETEROMULTIMER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Francis P. Bigley  
STREET: 126 E. Lincoln Avenue, P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,506  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/05768  
FILING DATE:  
APPLICATION NUMBER: US 08/242,811  
FILING DATE: 13-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bigley, Francis P.  
REGISTRATION NUMBER: 36,356  
REFERENCE/DOCKET NUMBER: 19226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-1249  
TELEFAX: (908) 594-4270  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2238 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-08-732-506-1

Query Match 11.8%; Score 15; DB 1; Length 2238;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AAGCTGAAAACAGC 71  
|||||  
Db 1106 AAGCTGAAAACAGC 1092

RESULT 44  
PCT-US95-05768-1/c  
Sequence 1, Application PC/TUS9505768  
GENERAL INFORMATION:  
APPLICANT: Kaczorowski, Gregory J.  
APPLICANT: Garcia, Maria L.  
APPLICANT: Leonard, Reid J.  
APPLICANT: McManus, Owen B.  
APPLICANT: Swanson, Richard J.  
APPLICANT: Folander, Kimberly L.  
TITLE OF INVENTION: NOVEL HETEROMULTIMER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Francis P. Bigley  
STREET: 126 E. Lincoln Avenue, P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/05768  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/242,811  
FILING DATE: 13-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bigley, Francis P.  
REGISTRATION NUMBER: 36,356  
REFERENCE/DOCKET NUMBER: 19226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-1249  
TELEFAX: (908) 594-4270

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2238 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
PCT-US95-05768-1

Query Match 11.8%; Score 15; DB 5; Length 2238;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AAGCTGAAAACAAGC 71  
|||||  
DB 1106 AAGCTGAAAACAAGC 1092

RESULT 45  
US-07-959-943-10/c  
Sequence 10, Application US/07959943  
Patent No. 5418162  
GENERAL INFORMATION:  
APPLICANT: Blakey, Randy D.  
APPLICANT: Fremieu Jr., Robert T.  
APPLICANT: Caron, Marc G.  
TITLE OF INVENTION: Serotonin Transporter cDNA  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley, Bell, Seltzer, Park  
ADDRESSEE: and  
ADDRESS: Gibson  
STREET: Post Office Drawer 31107  
CITY: Raleigh  
STATE: No. 5418162ch Carolina  
COUNTRY: U.S.A.  
ZIP: 27622  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/959,943  
FILING DATE: 19921014  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405.38a  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-881-3140  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
POSITION IN GENOME:  
UNITS: 2278 basepairs  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73..1962  
US-07-959-943-10

Query Match 11.8%; Score 15; DB 1; Length 2508;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 AAGGTCAGGTGATG 37  
|||||

DB 737 AAGGTCAGGTGATG 723

RESULT 46  
US-09-074-658-69/c  
Sequence 69, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quljun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2718 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-69

Query Match 11.8%; Score 15; DB 3; Length 2718;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TACTGTGCGCATCTG 21  
|||||  
DB 442 TACTGTGCGCATCTG 428

RESULT 47  
US-09-620-312D-552/c  
Sequence 552, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunging

APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhwei  
APPLICANT: John Tillinghaast  
APPLICANT: Dimanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
FILE REFERENCE: 784CIP28  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317.  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pc\_fl\_genes Version 1.0  
SEQ ID NO 552  
LENGTH: 3582  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (392)..(2647)  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(3582)  
OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-552

Query Match 11.8%; Score 15; DB 4; Length 3582;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 TGGGAGACTGTAC 106  
DB 3014 TGGGAGACTGTAC 3000

RESULT 48  
US-08-026-138E-17/c  
Sequence 17, Application US/08026138E  
Patent No. 5502166  
GENERAL INFORMATION:  
APPLICANT: Masayoshi MISHINA  
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nishiohata Residence 1-107  
STREET: 5214, Nishiohata-machi  
CITY: Niigata-shi  
STATE: Niigata-ken  
COUNTRY: JAPAN  
ZIP: 951  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS V.5  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/026,138E  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 39563/1992  
FILING DATE: 26-FEB-1992  
APPLICATION NUMBER: JP 173155/1992  
FILING DATE: 30-JUN-1992  
APPLICATION NUMBER: JP 215017/1992  
FILING DATE: 12-AUG-1992  
APPLICATION NUMBER: JP 303878/1992  
FILING DATE: 13-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hamburg, C. Bruce  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-4551

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 986-2340  
TELEFAX: (212) 953-7733  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4368 nucleic acids  
TYPE: nucleic acid  
STRANDEDNESS: double strand  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: mouse  
TISSUE TYPE: brain  
PUBLICATION INFORMATION:  
AUTHORS: Masayoshi MISHINA  
TITLE: NOVEL PROTEINS AND GENES CODING THE SAME  
RELEVANT RESIDUES IN SEQ ID NO: 17: FROM 1 TO 4368  
US-08-026-138E-17

Query Match 11.8%; Score 15; DB 1; Length 4368;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 TTGTGCTGGGAG 98  
DB 3281 TTGTGCTGGGAG 3267

RESULT 49  
US-08-026-138E-6/c  
Sequence 6, Application US/08026138E  
Patent No. 5502166  
GENERAL INFORMATION:  
APPLICANT: Masayoshi MISHINA  
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nishiohata Residence 1-107  
STREET: 5214, Nishiohata-machi  
CITY: Niigata-shi  
STATE: Niigata-ken  
COUNTRY: JAPAN  
ZIP: 951  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS V.5  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/026,138E  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 39563/1992  
FILING DATE: 26-FEB-1992  
APPLICATION NUMBER: JP 173155/1992  
FILING DATE: 30-JUN-1992  
APPLICATION NUMBER: JP 215017/1992  
FILING DATE: 12-AUG-1992  
APPLICATION NUMBER: JP 303878/1992  
FILING DATE: 13-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hamburg, C. Bruce  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-4551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 986-2340  
TELEFAX: (212) 953-7733  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4446 nucleic acids  
TYPE: nucleic acid  
STRANDEDNESS: double strand  
TOPOLOGY: linear



MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: mouse  
TISSUE TYPE: brain  
PUBLICATION INFORMATION:  
AUTHORS: Maayoshi MISHINA  
TITLE: NOVEL PROTEINS AND GENES CODING THE SAME  
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 to 4446  
US-08-026-138E-6

Query Match 11.8%; Score 15; DB 1; Length 4446;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 TTGTGCTGTGGGAG 98  
|||||  
Db 3359 TTGTGCTGTGGGAG 3345

RESULT 50  
US-09-215-694-20/c  
Sequence 20. Application US/09215694B  
Patent No. 6391583  
GENERAL INFORMATION:  
APPLICANT: Wisconsin Alumni Research Foundation  
APPLICANT: Hutchinson, Charles R.  
APPLICANT: Kennedy, Jonathan n.m.i  
APPLICANT: Park, Cheonsok n.m.i  
TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS  
FILE REFERENCE: 960296.95718  
CURRENT APPLICATION NUMBER: US/09/215,694B  
CURRENT FILING DATE: 1999-12-18  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 20  
LENGTH: 5053  
TYPE: DNA  
ORGANISM: Aspergillus terreus  
US-09-215-694-20

Query Match 11.8%; Score 15; DB 4; Length 5053;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CCAGTGATGATGC 42  
|||||  
Db 210 CCAGTGATGATGC 196

Search completed: November 7, 2003, 11:59:33  
Job time : 37.4064 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 11:58:22 ; Search time 489,785 Seconds  
(without alignments)  
827.407 Million cell updates/sec

Title: US-09-509-779-1\_COPY\_154\_280

Perfect score: 127  
Sequence: 1 GCGCATACCTGTGCATCT.....ATTCCCTGCACACTGCTGC 127

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2141354 seqs, 1595478879 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	25.2	342	9	US-09-826-312-7
2	32	25.2	342	14	US-10-108-767-7
3	32	25.2	342	14	US-10-152-156-7
4	32	25.2	342	10	US-09-796-692-8547
5	32	25.2	612	10	US-10-040-863-8547
6	32	25.2	612	14	US-10-040-863-8547
7	32	25.2	836	10	US-09-764-864-38
8	29	22.8	836	10	US-09-764-864-48
9	29	22.8	535	14	US-10-102-524-396
10	25	19.7	60	12	US-09-908-975-1285
11	25	16.5	933	14	US-10-198-846-7306
12	19	15.0	498	10	US-09-764-877-3347
13	19	15.0	3372	12	US-10-027-632-113318
14	19	15.0	3372	13	US-10-027-632-113318
15	18	14.2	441	12	US-10-027-632-118349
16	18	14.2	441	13	US-10-027-632-118349

17	18	14.2	833	12	US-10-027-632-162982	Sequence 162982.
18	18	14.2	833	12	US-10-027-632-162983	Sequence 162983.
19	18	14.2	833	13	US-10-027-632-162982	Sequence 162982.
20	18	14.2	833	13	US-10-027-632-162983	Sequence 162983.
21	18	14.2	930	12	US-10-027-632-11061	Sequence 31061, A
22	18	14.2	930	13	US-10-027-632-11061	Sequence 31061, A
23	17	13.4	252	12	US-10-027-632-276727	Sequence 276727.
24	17	13.4	252	13	US-10-027-632-276727	Sequence 276727.
25	17	13.4	348	10	US-09-878-574-2599	Sequence 2599, Ap
26	17	13.4	355	9	US-09-864-761-4974	Sequence 4974, Ap
27	17	13.4	355	9	US-09-864-761-21703	Sequence 21703, A
28	17	13.4	464	14	US-10-066-543-1656	Sequence 1656, Ap
29	17	13.4	481	10	US-09-878-574-1344	Sequence 1344, Ap
30	17	13.4	500	11	US-09-918-995-11855	Sequence 11855, A
31	17	13.4	575	12	US-10-027-632-224793	Sequence 224793.
32	17	13.4	575	13	US-10-027-632-224793	Sequence 224793.
33	17	13.4	712	12	US-10-027-632-145248	Sequence 145248.
34	17	13.4	712	13	US-10-027-632-145248	Sequence 145248.
35	17	13.4	1781	12	US-10-027-632-255720	Sequence 255720.
36	17	13.4	1781	13	US-10-027-632-255720	Sequence 255720.
37	17	13.4	2904	14	US-10-281-904-3	Sequence 3, Appli
38	17	13.4	3494	10	US-09-880-107-2295	Sequence 2295, Ap
39	17	13.4	3494	14	US-10-205-823-31	Sequence 31, Appl
40	17	13.4	3494	14	US-10-264-374-200	Sequence 200, App
41	17	13.4	3494	14	US-10-281-904-1	Sequence 1, Appli
42	17	13.4	3681	12	US-10-133-013-221	Sequence 221, App
43	17	13.4	3686	10	US-09-981-353-121	Sequence 121, App
44	17	13.4	3686	13	US-10-044-050-218	Sequence 218, App
45	17	13.4	3686	14	US-10-084-817-86	Sequence 86, Appl

## ALIGNMENTS

RESULT 1  
US-09-826-312-7  
; Sequence 7, Application US/09826312  
; Patent No. US20020042083A1  
; GENERAL INFORMATION:  
; APPLICANT: Issakani, Sarkiz D.  
; APPLICANT: Huang, Jiating  
; APPLICANT: Sheung, Julie  
; APPLICANT: Pray, Todd R.  
; TITLE OF INVENTION: UBIQUITIN LIGASE ASSAY  
; FILE REFERENCE: A-68613-1/RMS/JTD  
; CURRENT APPLICATION NUMBER: US/09/826.312  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 09/542.497  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 342  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-826-312-7

Query Match 25.2%; Score 32; DB 9; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCGATCTGACGGTCCAGGTGATGATGATGCTG 45  
DB 151 GCGATCTGACGGTCCAGGTGATGATGATGCTG 182

RESULT 2  
US-10-108-767-7  
; Sequence 7, Application US/10108767  
; Publication No. US20030104474A1  
; GENERAL INFORMATION:  
; APPLICANT: Issakani, Sarkiz D.  
; APPLICANT: Huang, Jiating

```

1  APPLICANT: Sheung, Julie
2  APPLICANT: Pray, Todd R.
3  TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENT
4  TITLE OF INVENTION: MODIFY THE ACTIVITY OF UBIQUITIN AGENTS
5  FILE REFERENCE: A-68613-5/RMS/DCF
6  CURRENT APPLICATION NUMBER: US/10/108,767
7  CURRENT FILING DATE: 2002-09-26
8  PRIOR APPLICATION NUMBER: US 09/542,497
9  PRIOR FILING DATE: 2000-04-03
10 PRIOR APPLICATION NUMBER: US 09/826,312
11 PRIOR FILING DATE: 2001-04-03
12 PRIOR APPLICATION NUMBER: US 10/091,139
13 PRIOR FILING DATE: 2002-03-04
14 NUMBER OF SEQ ID NOS: 27
15 SOFTWARE: PatentIn version 3.1
16 SEQ ID NO: 7
17 LENGTH: 342
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 US-10-108-767-7

```

Query Match 25.2%; Score 32; DB 14; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
US-10-152-156-7  
; Sequence 7, Application US/10152156  
; Publication No. US20030108947A1

```

/ TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENTS
/ TITLE OF INVENTION: MODIFY THE ACTIVITY OF UBIQUITIN AGENTS
/ FILE REFERENCE: A-68613-6/RMS/DCF
/ CURRENT APPLICATION NUMBER: US/10/152,156
/ CURRENT FILING DATE: 2002-05-20
/ PRIOR APPLICATION NUMBER: US 09/542,497
/ PRIOR FILING DATE: 2000-04-03
/ PRIOR APPLICATION NUMBER: US 09/826,312
/ PRIOR FILING DATE: 2001-04-03
/ PRIOR APPLICATION NUMBER: US 10/091,174
/ PRIOR FILING DATE: 2002-03-04
/ PRIOR APPLICATION NUMBER: US 10/091,139
/ PRIOR FILING DATE: 2002-03-04
/ PRIOR APPLICATION NUMBER: US 10/109,460
/ PRIOR FILING DATE: 2002-03-26
/ PRIOR APPLICATION NUMBER: US 10/108,767
/ PRIOR FILING DATE: 2002-03-26
/ PRIOR APPLICATION NUMBER: US 60/291,836
/ PRIOR FILING DATE: 2001-05-18
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 7
/ LENGTH: 342
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-152-156-7

```

Query Match	25.2%	Score 32;	DB 14;	Length 342;
Best Local Similarity	100.0%	Pred. No. 1.7e-07;		
Matches	32;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0
QY	14	GCCATCTGCAGGGTCCAGGTGATGATGCCCTG	45	
Db	151	GCCATCTGCAGGGTCCAGGTGATGATGCCCTG	182	

```

RESULT 4
US-09-796-692-8547
/ Sequence 8547, Application US/09796692
/ Publication No. US20020198362A1
/ GENERAL INFORMATION:
/ APPLICANT: Gaiger, Alexander
/ APPLICANT: Aigate, Paul A.
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
/ TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
/ FILE REFERENCE: 2077.001200
/ CURRENT APPLICATION NUMBER: US/09/796,692
/ CURRENT FILING DATE: 2001-03-01
/ PRIOR APPLICATION NUMBER: 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: 60/223,416
/ PRIOR FILING DATE: 2000-08-04
/ PRIOR APPLICATION NUMBER: 60/223,378
/ PRIOR FILING DATE: 2000-08-07
/ NUMBER OF SEQ ID NOS: 9597
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 8547
/ LENGTH: 612
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (485)
/ OTHER INFORMATION: n=A,T,C or G
US-09-796-692-8547

```

```

Query March 25.2%: Score 32; DB 10; Length 612;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
|||||
Db 35 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 66

RESULT 5
US-10-040-862-8547
; Sequence 8547, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US

```

```

; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8547
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (485)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-8547
```

```

Query Match      25.2%; Score 32; DB 14; Length 612;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
Db      35 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 66
```

```

RESULT 6
US-09-764-864-39
; Sequence 39, Application US/09/764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 836
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-39
```

```

Query Match      25.2%; Score 32; DB 10; Length 836;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
```

```

Db      166 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 197
```

```

RESULT 7
US-09-764-864-498
; Sequence 498, Application US/09/764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 498
; LENGTH: 836
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-498
```

```

Query Match      25.2%; Score 32; DB 10; Length 836;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      14 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 45
Db      219 GCCATCTGCAGGGTCCAGGTGATGATGCCTG 250
```

```

RESULT 8
US-10-102-524-396
; Sequence 396, Application US/10/102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 396
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 14, 24, 32, 206, 240, 361, 384, 439, 442
; OTHER INFORMATION: n = A,T,C or G
US-10-102-524-396
```

```

Query Match      22.8%; Score 29; DB 14; Length 535;
Best Local Similarity 100.0%; Pred. No. 7.1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      71 CAAGAGACTGTGTGTGTCTTGGGGAGA 99
Db      71 CAAGAGACTGTGTGTGTCTTGGGGAGA 99
```

```
RESULT 9
```

```
US-10-198-846-12407
; Sequence 12407, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12407
; LENGTH: 1024
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 1022, 1023, 1024
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12407

Query Match
Best Local Similarity 22.8%; Score 29; DB 14; Length 1024;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CAAGAGACTGTGTGTGTGTGTGGGAGA 99
DB 348 CAAGAGACTGTGTGTGTGTGTGGGAGA 376

RESULT 10
US-09-908-975-13285
; Sequence 13285, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: WASSERMAN, Avi
; APPLICANT: SHOSHAN, Avi
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Eli
; APPLICANT: FAIGLER, Simcha
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13285
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-13285

Query Match
Best Local Similarity 19.7%; Score 25; DB 12; Length 60;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GCCATTCGACGGTCCAGGTGATGG 38
DB 36 GCCATTCGACGGTCCAGGTGATGG 60
```

```
RESULT 11
US-10-198-846-7306
; Sequence 7306, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7306
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 36, 150, 168, 187, 224, 266, 305, 375, 378, 405, 429, 438,
; LOCATION: 439, 454, 458, 464, 472, 485, 491, 495, 516, 536, 537, 538,
; LOCATION: 544, 545, 548, 550, 558, 559, 563, 565, 567, 570, 578, 581,
; LOCATION: 585, 617, 633, 635, 636, 637, 641, 644, 647, 651, 657
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 672, 674, 680, 690, 692, 694, 695, 699, 704, 705, 707, 709,
; LOCATION: 725, 733, 740, 743, 744, 747, 748, 754, 756, 757, 762, 765,
; LOCATION: 769, 771, 774, 775, 777, 783, 784, 786, 788, 792, 795, 797,
; LOCATION: 798, 801, 802, 805, 809, 812, 813, 814, 815, 817, 818
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 819, 824, 827, 836, 838, 846, 851, 852, 855, 860, 865, 871,
; LOCATION: 876, 882, 893, 901, 903, 904, 905, 908, 909, 910, 913, 915,
; LOCATION: 919, 925, 927, 930, 931
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7306

Query Match
Best Local Similarity 16.5%; Score 21; DB 14; Length 933;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 CATTCTTCACACACTGCTGC 127
DB 93 CATTCTTCACACACTGCTGC 113

RESULT 12
US-09-764-877-3347
; Sequence 3347, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3347
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3347
```

Query Match 15.0%; Score 19; DB 10; Length 498;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AACCAAGCAGAGGACTGT 82  
DB 403 AACCAAGCAGAGGACTGT 421

## RESULT 13

US-10-027-632-113318/c  
; Sequence 113318, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 113318  
; LENGTH: 3372  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-113318

Query Match 15.0%; Score 19; DB 12; Length 3372;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 AGAGACTGTGTGTGTC 91  
DB 1339 AGAGACTGTGTGTGTC 1321

## RESULT 14

US-10-027-632-113318/c  
; Sequence 113318, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 113318  
; LENGTH: 3372  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-113318

Query Match 15.0%; Score 19; DB 13; Length 3372;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 AGAGACTGTGTGTGTC 91  
DB 1339 AGAGACTGTGTGTGTC 1321

## RESULT 15

US-10-027-632-113349  
; Sequence 113349, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 113349  
; LENGTH: 441  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-113349

Query Match 14.2%; Score 18; DB 12; Length 441;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TCTGGGAGAGGTGAAC 107  
DB 3 TCTGGGAGAGGTGAAC 20

## RESULT 16

US-10-027-632-113349  
; Sequence 113349, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12

```
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 138349
;; LENGTH: 441
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-138349
```

```
Query Match      14.2%: Score 18; DB 13; Length 441;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      90 TCTGGGAGAGTGTACC 107
      |||||
Db      3 TCTGGGAGAGTGTACC 20
```

```
RESULT 17
US-10-027-632-162982
;; Sequence 162982, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 162982
;; LENGTH: 833
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-162982
```

```
Query Match      14.2%: Score 18; DB 12; Length 833;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      53 TGTCAAGCTGAAACAG 70
      |||||
Db      194 TGTCAAGCTGAAACAG 211
```

```
RESULT 18
US-10-027-632-162983
```

```
;; Sequence 162983, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 162983
;; LENGTH: 833
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-162983
```

```
Query Match      14.2%: Score 18; DB 12; Length 833;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      53 TGTCAAGCTGAAACAG 70
      |||||
Db      194 TGTCAAGCTGAAACAG 211
```

```
RESULT 19
US-10-027-632-162982
;; Sequence 162982, Application US/10027632
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 162982
;; LENGTH: 833
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-162982
```

```
Query Match      14.2%: Score 18; DB 13; Length 833;
```



Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 TGTCAAGCTGAACAAG 70  
DB 194 TGTCAAGCTGAACAAG 211

## RESULT 20

US-10-027-632-162983  
Sequence 162983, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 162983  
LENGTH: 833  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-162983

Query Match 14.2%; Score 18; DB 13; Length 833;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 TGTCAAGCTGAACAAG 70  
DB 194 TGTCAAGCTGAACAAG 211

## RESULT 21

US-10-027-632-31061  
Sequence 31061, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31061  
LENGTH: 930  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-31061

Query Match 14.2%; Score 18; DB 12; Length 930;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ATTCTTCACACACGTCT 125  
DB 774 ATTCTTCACACACGTCT 791

## RESULT 22

US-10-027-632-31061  
Sequence 31061, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31061  
LENGTH: 930  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-31061

Query Match 14.2%; Score 18; DB 13; Length 930;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ATTCTTCACACACGTCT 125  
DB 774 ATTCTTCACACACGTCT 791

## RESULT 23

US-10-027-632-276727/c  
Sequence 276727, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676

```

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 276727
; LENGTH: 252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-276727
```

```
Query Match          13.4% Score 17; DB 12; Length 252;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      21 GCAGGCTCCAGTGATG 37
      |||||
DB      51 GCAGGCTCCAGTGATG 35.
```

```
RESULT 24
US-10-027-632-276727/c
; Sequence 276727, Application US/10027632
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 276727
; LENGTH: 252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-276727
```

```
Query Match          13.4% Score 17; DB 13; Length 252;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      21 GCAGGCTCCAGTGATG 37
      |||||
DB      51 GCAGGCTCCAGTGATG 35
```

```
RESULT 25
US-09-878-574-2599/c
; Sequence 2599, Application US/09878574
; Patent No. US20020110548A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO: 2559
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: L1B3028-023-Q1-B1-B1
US-09-878-574-2599
```

```
Query Match          13.4% Score 17; DB 10; Length 348;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      101 TGTAACTTCCTTCCA 117
      |||||
DB      205 TGTAACTTCCTTCCA 189
```

```
RESULT 26
US-09-864-761-4974
; Sequence 4974, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
```

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4974
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031056.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; US-09-864-761-4974

Query Match      13.4%; Score 17; DB 9; Length 355;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      25 GGTCCAGGTGATGATG 41
Db      281 GGTCCAGGTGATGATG 297

RESULT 27
US-09-864-761-21703
; Sequence 21703, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21703
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031056.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: NT HIT: AB019786.1, EVALUATE 1.20e-02
; OTHER INFORMATION: EST_HUMAN HIT: AA495851.1, EVALUATE 2.20e-01
; US-09-864-761-21703

Query Match      13.4%; Score 17; DB 9; Length 355;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      25 GGTCCAGGTGATGATG 41
Db      281 GGTCCAGGTGATGATG 297

RESULT 28
US-10-066-543-1656
; Sequence 1656, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1656
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 211, 453, 454
; OTHER INFORMATION: n = A,T,C or G

```

US-10-066-543-1656

Query Match 13.4%; Score 17; DB 14; Length 464;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGAGGGTCCAG 31  
|||||  
DB 433 CCATCTGAGGGTCCAG 449

RESULT 29  
US-09-878-574-1344/C

; Sequence 1344; Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 1344  
; LENGTH: 481  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: LIB3028-039-Q1-B1-E10  
US-09-878-574-1344

Query Match 13.4%; Score 17; DB 10; Length 481;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 TGTACCACTTCTTCCA 117  
|||||  
DB 439 TGTACCACTTCTTCCA 423

RESULT 30  
US-09-918-995-31855/C

; Sequence 31855; Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31855  
; LENGTH: 500  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(500)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31855

Query Match 13.4%; Score 17; DB 11; Length 500;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGAGGGTCCAG 31

DB 482 CCATCTGAGGGTCCAG 466

RESULT 31  
US-10-027-632-224793

; Sequence 224793; Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 224793  
; LENGTH: 575  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-224793

Query Match 13.4%; Score 17; DB 12; Length 575;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAACACAGCA 72  
|||||  
DB 46 CAAGCTGAACACAGCA 62

RESULT 32  
US-10-027-632-224793

; Sequence 224793; Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 224793

LENGTH: 575  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-224793

Query Match 13.4%; Score 17; DB 13; Length 575;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAAGCTGAAACAGCA 72  
DB 46 CAAGCTGAAACAGCA 62

## RESULT 33

US-10-027-632-145248/C  
Sequence 145248, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 145248  
LENGTH: 712  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-145248

Query Match 13.4%; Score 17; DB 12; Length 712;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ATCTGAGGTCACAGT 33  
DB 689 ATCTGAGGTCACAGT 673

## RESULT 34

US-10-027-632-145248/C  
Sequence 145248, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 145248  
LENGTH: 712  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-145248

Query Match 13.4%; Score 17; DB 13; Length 712;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ATCTGAGGTCACAGT 33  
DB 689 ATCTGAGGTCACAGT 673

## RESULT 35

US-10-027-632-255720  
Sequence 255720, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 255720  
LENGTH: 1781  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-255720

Query Match 13.4%; Score 17; DB 12; Length 1781;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 TTCCTCCACACAGTCT 125  
DB 1695 TTCCTCCACACAGTCT 1711

## RESULT 36

US-10-027-632-255720  
Sequence 255720, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome

```
FILE REFERENCE: 108827,129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 255720
LENGTH: 1781
TYPE: DNA
ORGANISM: Human
US-10-027-632-255720
```

```
Query Match 13.4% Score 17; DB 13; Length 1781;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 109 TTCCTTCACACTGCT 125
Db 1695 TTCCTTCACACTGCT 1711
```

```
RESULT 37
US-10-281-904-3/c
Sequence 3, Application US/10281904
Publication No. US20030119036A1
GENERAL INFORMATION:
APPLICANT: Chun, Miyoung
TITLE OF INVENTION: METHODS OF USING 48149, A HUMAN
TITLE OF INVENTION: AMINOPEPTIDASE FAMILY MEMBER
FILE REFERENCE: MP101-174P1RM
CURRENT APPLICATION NUMBER: US/10/281,904
CURRENT FILING DATE: 2002-10-28
PRIOR APPLICATION NUMBER: 60/335,084
PRIOR FILING DATE: 2001-10-31
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 2904
TYPE: DNA
ORGANISM: Homo sapien
US-10-281-904-3
```

```
Query Match 13.4% Score 17; DB 14; Length 2904;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 15 CCATCTGCAGGCTCCAG 31
Db 1642 CCATCTGCAGGCTCCAG 1626
```

```
RESULT 38
US-09-880-107-2295/c
Sequence 2295, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
```

```
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 2295
LENGTH: 3494
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 M22324
US-09-880-107-2295
```

```
Query Match 13.4% Score 17; DB 10; Length 3494;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 15 CCATCTGCAGGCTCCAG 31
Db 1762 CCATCTGCAGGCTCCAG 1746
```

```
RESULT 39
US-10-205-823-31/c
Sequence 31, Application US/10205823
Publication No. US20030108963A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula
APPLICANT: Gorbacheva, Bella
APPLICANT: Hoersch, Sebastian
APPLICANT: Kamatkari, Shubhangi
APPLICANT: Monsey, Angela M.
APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/307,982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314,356
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/325,020
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 31
LENGTH: 3494
TYPE: DNA
ORGANISM: Homo sapiens
US-10-205-823-31
```

```
Query Match 13.4% Score 17; DB 14; Length 3494;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 15 CCATCTGCAGGCTCCAG 31
```

Db 1762 CCATCTGACGGGTCAG 1746

RESULT 40  
US-10-264-374-200/c  
; Sequence 200, Application US/10264374  
; Publication No. US20030113320A1  
; GENERAL INFORMATION:  
; APPLICANT: Rusolanti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: P-LJ 3203  
; CURRENT APPLICATION NUMBER: US/10/264,374  
; CURRENT FILING DATE: 2002-10-03  
; PRIOR APPLICATION NUMBER: US/09/139,802  
; PRIOR FILING DATE: 1998-08-25  
; PRIOR APPLICATION NUMBER: 08/926,914  
; PRIOR FILING DATE: 1997-09-10  
; PRIOR APPLICATION NUMBER: 08/710,067  
; PRIOR FILING DATE: 1996-09-10  
; NUMBER OF SEQ ID NOS: 226  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 200  
; LENGTH: 3494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (121)..(3024)  
US-10-264-374-200

Query Match 13.4%; Score 17; DB 14; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCAG 31  
Db 1762 CCATCTGACGGGTCAG 1746

RESULT 41  
US-10-281-904-1/c  
; Sequence 1, Application US/10281904  
; Publication No. US20030119036A1  
; GENERAL INFORMATION:  
; APPLICANT: Chun, Miyoung  
; TITLE OF INVENTION: METHODS OF USING 48149, A HUMAN  
; TITLE OF INVENTION: AMINOPEPTIDASE FAMILY MEMBER  
; FILE REFERENCE: MP101-174P1RM  
; CURRENT APPLICATION NUMBER: US/10/281,904  
; CURRENT FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: 60/335,084  
; PRIOR FILING DATE: 2001-10-31  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-281-904-1

Query Match 13.4%; Score 17; DB 14; Length 3494;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCAG 31  
Db 1762 CCATCTGACGGGTCAG 1746

RESULT 42

US-10-133-013-221/c  
; Sequence 221, Application US/10133013  
; Publication No. US20030166903A1  
; GENERAL INFORMATION:  
; APPLICANT: Astromoff, Anna  
; APPLICANT: Bandman, Olga  
; APPLICANT: Cocks, Benjamin G.  
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE  
; FILE REFERENCE: PA-0049 US  
; CURRENT APPLICATION NUMBER: US/10/133,013  
; CURRENT FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 60/287,067  
; PRIOR FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 271  
; SOFTWARE: PERL Program  
; SEQ ID NO 221  
; LENGTH: 3681  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20030166903A1 1136923.19  
US-10-133-013-221

Query Match 13.4%; Score 17; DB 12; Length 3681;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCAG 31  
Db 1960 CCATCTGACGGGTCAG 1944

RESULT 43  
US-09-981-353-121/c  
; Sequence 121, Application US/09981353  
; Patent No. US20020160382A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 121  
; LENGTH: 3686  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1344279CB1  
US-09-981-353-121

Query Match 13.4%; Score 17; DB 10; Length 3686;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCAG 31  
Db 1962 CCATCTGACGGGTCAG 1946

RESULT 44  
US-10-044-090-218/c  
; Sequence 218, Application US/10044090  
; Publication No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090

```

; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 218
; LENGTH: 3686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1344279CB1
US-10-044-090-218
```

```

Query Match          13.4%; Score 17; DB 13; Length 3686;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      15 CCATCTGCAGGGTCCAG 31
DB      1962 CCATCTGCAGGGTCCAG 1946
```

```

RESULT 45
US-10-084-817-86/c
; Sequence 86, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Pion
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 86
; LENGTH: 3686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 1344279CB1
US-10-084-817-86
```

```

Query Match          13.4%; Score 17; DB 14; Length 3686;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      15 CCATCTGCAGGGTCCAG 31
DB      1962 CCATCTGCAGGGTCCAG 1946
```

```

RESULT 46
US-09-925-297-338/c
; Sequence 338, Application US/09925297
; Patent No. US2002081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 338
```

```

; LENGTH: 3728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3707)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3713)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-338
```

```

Query Match          13.4%; Score 17; DB 9; Length 3728;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      15 CCATCTGCAGGGTCCAG 31
DB      1972 CCATCTGCAGGGTCCAG 1956
```

```

RESULT 47
US-10-133-013-222/c
; Sequence 222, Application US/10133013
; Publication No. US20030166903A1
; GENERAL INFORMATION:
; APPLICANT: Astromoff, Anna
; APPLICANT: Bandman, Olga
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE
; FILE REFERENCE: PA-0049 US
; CURRENT APPLICATION NUMBER: US/10/133,013
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/287,067
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 271
; SOFTWARE: PERL Program
; SEQ ID NO 222
; LENGTH: 3815
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030166903A1 1136923.18
US-10-133-013-222
```

```

Query Match          13.4%; Score 17; DB 12; Length 3815;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      15 CCATCTGCAGGGTCCAG 31
DB      1961 CCATCTGCAGGGTCCAG 1945
```

```

RESULT 48
US-09-764-891-6245/c
; Sequence 6245, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAMM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6245
; LENGTH: 11150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6245
```



Query Match 13.4%; Score 17; DB 11; Length 11150;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCCAG 31  
|||||  
Db 76 CCATCTGACGGGTCCAG 60

RESULT 49

US-10-205-428-561/c  
; Sequence 561, Application US/10205428  
; Publication No. US20030108907A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P117C1  
; CURRENT APPLICATION NUMBER: US/10/205,428  
; PRIOR FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1019  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 561  
; LENGTH: 11150  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-205-428-561

Query Match 13.4%; Score 17; DB 14; Length 11150;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATCTGACGGGTCCAG 31  
|||||  
Db 76 CCATCTGACGGGTCCAG 60

RESULT 50

US-09-795-668-1  
; Sequence 1, Application US/09795668  
; Patent No. US20020045577A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinhorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345,2004-001  
; CURRENT APPLICATION NUMBER: US/09/795,668  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1503841

; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: r=g or a  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: y=c/u or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: m=a or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: b=g or c or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: d=a or g or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: h=a or c or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: v=a or g or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: n=a or g or c or t/u  
US-09-795-668-1

Query Match 13.4%; Score 17; DB 9; Length 1503841;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 CTGAAACACGACAGAG 76  
|||||  
Db 1166621 CTGAAACACGACAGAG 1166637

Search completed: November 7, 2003, 13:55:39  
Job time : 491.785 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 02:14:16 ; Search time 1294.79 Seconds  
(without alignments)  
2383.909 Million cell updates/sec

Title: US-09-509-779-1\_COPY\_154\_280

Perfect score: 127  
Sequence: 1 GGGCATTACCTGTGCATCT.....ATTCTTCACAACTGCTGC 127

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estcov:\*  
5: em\_estcov:\*  
6: em\_estcov:\*  
7: em\_estcov:\*  
8: em\_estcov:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: em\_esthum:\*  
18: em\_esthum:\*  
19: em\_esthum:\*  
20: em\_esthum:\*  
21: em\_esthum:\*  
22: em\_esthum:\*  
23: em\_esthum:\*  
24: em\_esthum:\*  
25: em\_esthum:\*  
26: em\_esthum:\*  
27: em\_esthum:\*  
28: gb\_est1:\*  
29: gb\_est2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	263	10	BE624166
2	127	100.0	291	14	CB570887
3	127	100.0	296	9	AA231201
4	127	100.0	303	10	BE624155

5	127	100.0	309	9	AA000941
6	127	100.0	310	13	BY344528
7	127	100.0	319	13	BY329058
8	127	100.0	321	14	W78263
9	127	100.0	322	13	BY125373
10	127	100.0	323	10	BE692680
11	127	100.0	325	10	BF012987
12	127	100.0	328	9	AA623723
13	127	100.0	328	13	BY342015
14	127	100.0	329	13	BY114710
15	127	100.0	332	9	AA110259
16	127	100.0	346	13	BY337380
17	127	100.0	348	13	BY106705
18	127	100.0	349	13	BY116440
19	127	100.0	351	13	BY170257
20	127	100.0	352	13	BY091208
21	127	100.0	355	10	BE089348
22	127	100.0	357	13	BY090965
23	127	100.0	361	13	BY302366
24	127	100.0	366	9	AA403365
25	127	100.0	367	9	AA144182
26	127	100.0	368	13	BY089817
27	127	100.0	371	13	BY091920
28	127	100.0	371	13	BY169149
29	127	100.0	376	13	BY172672
30	127	100.0	382	13	BY161786
31	127	100.0	392	13	BY307739
32	127	100.0	393	14	CD565301
33	127	100.0	394	13	BY278099
34	127	100.0	395	13	BY095585
35	127	100.0	397	10	BE277656
36	127	100.0	401	10	BF011540
37	127	100.0	409	14	W34101
38	127	100.0	414	9	AA499298
39	127	100.0	416	13	BE469845
40	127	100.0	418	13	BY238589
41	127	100.0	427	10	BE333776
42	127	100.0	429	10	BE692461
43	127	100.0	440	9	AA815479
44	127	100.0	443	9	AA239898
45	127	100.0	445	14	W34374

## ALIGNMENTS

RESULT 1  
BE624166  
LOCUS  
DEFINITION  
IMAGE:3374593 5' similar to WP:R10A10.2 CE12670 ZINC FINGER PROTEIN  
, mRNA sequence.

ACCESSION  
BE624166  
VERSION  
BE624166.1 GI:9904582

SOURCE  
Mus musculus (house mouse)

REFERENCE  
1 (bases 1 to 263)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
National Cancer Institute, Cancer Genome Anatomy Project (CCAP),  
Tumor Gene Index

JOURNAL  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov

IMAGE Consortium (info@image.llnl.gov) for further information.  
MG:1084157

FEATURES  
Seq primer: -40RP from Gibco  
High quality sequence stop: 110.  
Location/Qualifiers  
1..263

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:3374593"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/clone_lib="Soares mammary gland NMLMG"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia) digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT733 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

```

## BASE COUNT

```
64 a 61 c 77 g 61 t
```

```

Query Match      100.0%; Score 127; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.7e-55;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTGGATACCTGTGGCATCTGCAGGCTCCAGATGATGATGCTGCTTCGATGCAAGC 60
DB 62 GTGGATACCTGTGGCATCTGCAGGCTCCAGATGATGATGCTGCTTCGATGCAAGC 121
QY 61 TGAACAAGCAAGAGAGACTGTGTGTGTGTGTGGGAGAGTAAACCATTCCTCCACA 120
DB 122 TGAACAAGCAAGAGAGACTGTGTGTGTGTGTGGGAGAGTAAACCATTCCTCCACA 181
QY 121 CTGCTGC 127
DB 182 CTGCTGC 188

```

```

RESULT 2
LOCUS CB570887 291 bp mRNA linear EST 02-APR-2003
DEFINITION AGENCOURT.13009672 NIH_MGC.165 Mus musculus cDNA clone
IMAGE:30276894 5', mRNA sequence.
ACCESSION CB570887
VERSION CB570887.1 GI:29490417
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Leslie L. Heckert
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: NDCM28 row: n column: 07
High quality sequence stop: 291.
Location/Qualifiers
1..291

```

## FEATURES

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30276894"
/tissue_type="primary cultures of Sertoli cells"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_165"
/notes="Organ: testis; Vector: pDNR-LIB; Site_1: Sfil

```

```

(gggccatggcc); Site 2: Sfil (ggcgccctggcc); 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCATTAAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCGCCGACACG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4 kb
(range 0.6-3.5 kb). 15/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH_MGC Library."

```

## BASE COUNT

```
61 a 70 c 97 g 62 t 1 others
```

```

Query Match      100.0%; Score 127; DB 14; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.7e-55;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTGGATACCTGTGGCATCTGCAGGCTCCAGATGATGATGCTGCTTCGATGCAAGC 60
DB 146 GTGGATACCTGTGGCATCTGCAGGCTCCAGATGATGATGCTGCTTCGATGCAAGC 205
QY 61 TGAACAAGCAAGAGAGACTGTGTGTGTGTGTGGGAGAGTAAACCATTCCTCCACA 120
DB 206 TGAACAAGCAAGAGAGACTGTGTGTGTGTGTGGGAGAGTAAACCATTCCTCCACA 265
QY 121 CTGCTGC 127
DB 266 CTGCTGC 272

```

```

RESULT 3
LOCUS AA231201 296 bp mRNA linear EST 26-FEB-1997
DEFINITION mw38g11.1 Soares mouse 3NME12 5' Mus musculus cDNA clone
IMAGE:673028 5', similar to WP:ZK287.5 CE06614 ;, mRNA sequence.
ACCESSION AA231201
VERSION AA231201.1 GI:1853558
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 296)
JOURNAL Geisel, S., Kucaba, T., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMNI Mouse EST Project
Unpublished
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LMN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:412732
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 258.
Location/Qualifiers
1..296

```

## FEATURES

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:673028"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse 3NME12 5"

```

BASE COUNT	62 a	74 c	92 g	67 t	1 others
ORIGIN					
Query Match	100.0%;	Score 127;	DB 9;	Length 296;	
Best Local Similarity	100.0%;	Pred. No. 1,7e-55;			
Matches 127;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTGGCATCTCTGTGGCATCTGTGAGGGTCCAGGTGATGGATCGCCCTTGATGTCAGC	60		
Db	109	GTGGCATCTCTGTGGCATCTGTGAGGGTCCAGGTGATGGATCGCCCTTGATGTCAGC	168		
QY	61	TGAACAAGCAGAGAGACTGTGTGTGTGTGTGGAGAGTGAACCATTCCTCCACAA	120		
Db	169	TGAACAAGCAGAGAGACTGTGTGTGTGTGTGGAGAGTGAACCATTCCTCCACAA	228		
QY	121	CTGCTGCG	127		
Db	229	CTGCTGCG	235		
RESULT 4					
BE624155					
LOCUS					
DEFINITION	BE624155 303 bp mRNA linear EST 24-AUG-2000				
	u42c01.y1 Soares mammary gland NMLMG Mus musculus cDNA clone				
	IMAGE:3374592.5' similar to WP:R10A10.2 CE12670 ZINC FINGER PROTEIN				
	; mRNA sequence.				
ACCESSION	BE624155				
VERSION	BE624155.1 GI:9904571				
KEYWORDS	EST.				
SOURCE	Mus musculus				
ORGANISM	Mus musculus (house mouse)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 303)				
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-rt@mail.nih.gov				
	This clone is available royalty-free through LNL; contact the				
	IMAGE Consortium (info@image.llnl.gov) for further information.				
	MG1:1084196				
	Seq primer: -40RP from Gibco				
	High quality sequence stop: 292.				
FEATURES					
source	location/Qualifiers				
	1..303				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:3374592"				
	/sex="female (lactating)"				
	/tissue_type="mammary gland"				
	/lab_host="DH10B"				
	/clone_idb="Soares mammary gland_NMLMG"				
	/note="Vector: pRT73-Pec (Pharmacia) with a modified				
	polylinker; 1st strand cDNA was prepared from mammary				
	gland tissue from a lactating female, and was then primed				
	with a Not I - oligo(dT) primer. Double-stranded cDNA was				
	ligated to Eco RI adaptors (Pharmacia), digested with Not				
	I and cloned into the Not I and Eco RI sites of the				
	modified pRT73 vector. Library is normalized. Library				
	was constructed by Bento Soares and M. Fatima Bonaldo. "				

[illegible]









Db 193 TGAACAAGCAGAGGACTGTTGCTGGGAGAGTGAACATTCCTCCACAA 252  
QY 121 CTGCTGC 127  
Db 253 CTGCTGC 259

RESULT 10  
LOCUS BE692680 323 bp mRNA linear EST 11-SEP-2000  
DEFINITION uw44b12.y1 Soares mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:3464831.5' similar to WP:R10A10.2 CE12670 ZINC FINGER PROTEIN  
; , mRNA sequence.  
ACCESSION BE692680  
VERSION BE692680.1 GI:10080304  
KEYWORDS EST  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 323)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Other ESTs: uw44b12.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1384191  
Seq primer: -40RP from Gibco  
High quality sequence stop: 260.

FEATURES  
source location/Qualifiers  
1..323  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3464831"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pRT73 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 73 a 80 c 102 g 68 t  
ORIGIN  
Query Match 100.0%; Score 127; DB 10; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1,7e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCATATCTGTGCATCTGCAGGTCAGGTGATGATGCTTCCTGATGTCAGC 60  
Db 125 GTGCATATCTGTGCATCTGCAGGTCAGGTGATGATGCTTCCTGATGTCAGC 184  
QY 61 TGAACAAGCAGAGGAGCTGTTGCTGGGAGAGTGAACATTCCTCCACAA 120  
Db 185 TGAACAAGCAGAGGAGCTGTTGCTGGGAGAGTGAACATTCCTCCACAA 244  
QY 121 CTGCTGC 127  
Db 245 CTGCTGC 251

RESULT 11  
BF012987

LOCUS BF012987 325 bp mRNA linear EST 29-DEC-2000  
DEFINITION ux83b03.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA  
clone IMAGE:3655085.5' similar to TR:Q9WTZ1 Q9WTZ1 ZINC RING FINGER  
PROTEIN SAG. ; , mRNA sequence.  
ACCESSION BF012987  
VERSION BF012987.1 GI:10713262  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 325)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schuck,R., Ritter  
B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1417389  
Seq primer: Primer name ambiguous  
High quality sequence stop: 316.

FEATURES  
source location/Qualifiers  
1..325  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD-1"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3655085"  
/sex="male"  
/tissue\_type="type B spermatogonia, pooled from multiple  
mice"  
/dev\_stage="8 day"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="McCarrey Eddy type B spermatogonia"  
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene  
); Site 1: XhoI; Site 2: EcoRI; cDNA oligo dt-primed  
[5'-(GA)10-ACTGATCGAGTGTGTTTGT-3'] and directionally  
cloned using 5' linkers 5'-AATCGCAGCAG-3' and  
5'-CTCGGCGG-3'. Size selection of >400bp material gives  
average insert size ranging from 1-2 kb. Library was mass  
excised (from lambda-uniZAP-XR) and resulting  
single-stranded phagemids were prepped and transformed  
into DH10B. Library contains 96% recombinants.  
References: J. Androl. 20:635-639 and Gene 25:263-269.  
Library constructed and donated by J. McCarrey, Ph.D.  
(Southwest Foundation for Biomedical Research, Dept. of  
Genetics); excision done by E.W. Eddy, Ph.D. (National  
Institutes of Health, National Institute of Environmental  
Health Sciences). Original lambda-based library is  
available through ATCC, catalog #63417."  
BASE COUNT 72 a 77 c 106 g 70 t  
ORIGIN  
Query Match 100.0%; Score 127; DB 10; Length 325;  
Best Local Similarity 100.0%; Pred. No. 1,7e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCATATCTGTGCATCTGCAGGTCAGGTGATGATGCTTCCTGATGTCAGC 60  
Db 81 GTGCATATCTGTGCATCTGCAGGTCAGGTGATGATGCTTCCTGATGTCAGC 140  
QY 61 TGAACAAGCAGAGGAGCTGTTGCTGGGAGAGTGAACATTCCTCCACAA 120  
Db 141 TGAACAAGCAGAGGAGCTGTTGCTGGGAGAGTGAACATTCCTCCACAA 200

QY 121 CTGCTGC 127  
 Db 201 CTGCTGC 207

RESULT 12  
 LOCUS 328 bp mRNA linear EST 14-OCT-1997  
 DEFINITION vq72g03.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone  
 IMAGE:1107892 5' similar to WP:ZK287.5 CE06614 ;, mRNA sequence.  
 AA623723  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 328)  
 REFERENCE  
 AUTHORS  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Gessel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished  
 JOURNAL  
 COMMENT  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LML ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:606060  
 Possible reversed clone; similarity on wrong strand  
 High quality sequence stop: 263.  
 Location/Qualifiers  
 1. 328  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J x DBA/2J F1"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1107892"  
 /tissue\_type="embryo"  
 /dev\_stage="2-cell"  
 /lab\_host="DH10B"  
 /clone\_lib="Knowles Solter mouse 2 cell"  
 /note="Organ: embryo; Vector: pBluescribe (modified);  
 Site\_1: MluI; Site\_2: SalI; Cloned unidirectionally from  
 mRNA prepared from 13,500 2-cell stage embryos. Primer:  
 SalI (dr): 5'-CGGTCAGCCGTCGACGCTTTTCTTTTCTT-3'.  
 CDNA  
 were cloned into the MluI/SalI sites of a modified  
 pBluescribe vector using commercial linkers (NBB).  
 Average insert size: 1.2 kb."  
 BASE COUNT  
 ORIGIN  
 76 a 78 c 104 g 70 t

Query Match 100.0%; Score 127; DB 9; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-55;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCATACCTGTCATCTGCAAGGTCAGGTGATGATGCTTCCTTCATGTCACAGC 60  
 Db 116 GTTCGATCTCTGTCATCTGCAAGGTCAGGTGATGATGCTTCCTTCATGTCACAGC 175  
 QY 61 TGAAGACAGCAAGAGAGCTGTGTGTGTGTGGGAGAGAGTACCACTTCCTTCCACAA 120  
 Db 176 TGAAGACAGCAAGAGAGCTGTGTGTGTGTGGGAGAGAGTACCACTTCCTTCCACAA 235  
 QY 121 CTGCTGC 127  
 Db 116 GTTCGATCTCTGTCATCTGCAAGGTCAGGTGATGATGCTTCCTTCATGTCACAGC 175

Db 236 CTGCTGC 242

RESULT 13  
 BY342015  
 LOCUS 328 bp mRNA linear EST 12-DEC-2002  
 DEFINITION BY342015 RIKEN full-length enriched, whole joints Mus musculus cDNA  
 clone l230037K21 5', mRNA sequence.  
 BY342015  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 328)  
 REFERENCE  
 AUTHORS  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nishida, I., Osato, N., Saito, R., Suzuki, H., Yamataka, T., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,  
 Quackenbush, J., Schmitt, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbett,  
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perce, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,  
 Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, D., Ring,  
 B.Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempel, C.A., Setou,  
 M., Shimada, K., Sultana, R., Takemata, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Whalley, C., Wang, Y.,  
 Watanabe, Y., Wells, L., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,  
 M., Yang, L., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komno, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 22354683  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Saitama-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hirozane,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Komno, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multichannel sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)



RESULT 15  
 A110259 332 bp mRNA linear EST 02-OCT-1997  
 LOCUS mp07g09.r1 Life Tech mouse embryo 8 5dpc 10664019 Mus musculus CDNA  
 DEFINITION clone IMAGE:568576 5' similar to WP:2K287.5 CE06614 ; mRNA  
 sequence.  
 ACCESSION A110259  
 VERSION A110259.1 GI:1662195  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 332)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI: 343224  
 Seq primer: -28M13 rev1 from Amersham  
 High quality sequence stop: 329.  
 Location/Qualifiers  
 1..332  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:568576"  
 /tissue\_type="embryo"  
 /dev\_stage="8.5dpc embryos"  
 /lab\_host="DH10B"  
 /clone\_lib="Life Tech mouse embryo 8 5dpc 10664019"  
 /note="Organ: whole embryo; Vector: PCMV-SPORT2; Site: 1:  
 Salt; Site 2: Not; Cloned unidirectionally. Primer:  
 Oligo dt. 8.5dpc embryos. PCMV-SPORT2 vector."  
 74 a 80 c 108 g 70 t  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 127; DB 9; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-55;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GTGCGATCCCTGTCATCTGCAGGTCAGGTGATGATGCTTCCTTCATGTCAGGC 60  
 Db 93 GTGCGATCCCTGTCATCTGCAGGTCAGGTGATGATGCTTCCTTCATGTCAGGC 152  
 Oy 61 TGAAGAAAG 120  
 Db 153 TGAAGAAAG 212  
 Oy 121 CTGCTGC 127  
 Db 213 CTGCTGC 219

RESULT 16  
 LOCUS BY337380 346 bp mRNA linear EST 11-DEC-2002  
 DEFINITION BY337380 RIKEN full-length enriched, whole joints Mus musculus CDNA  
 clone l230006P10 5', mRNA sequence.

ACCESSION BY337380  
 VERSION BY337380.1 GI:26532986  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 346)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,  
 Quackenbush, J., Schmitt, L. M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K. M., Blake, J. A., Bradt, D., Bruscia, V., Chochia, C., Corbett,  
 L. E., Cousins, S., Dalla, E., Dreagan, T. A., Fletcher, C. F., Forrest,  
 A., Fraser, K. S., Gasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,  
 Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M.,  
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,  
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Petrea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,  
 Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
 M. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
 R. D., Tomita, M., Verardo, R., Wagner, L., Walstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sakai, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsr.riken.go.jp,  
 URL: http://genome-gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arakawa, T.,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Numata, K., Ohno, M., Sakai, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multipillar sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Vasilius Aidinis ( Biomedical Sciences  
 Research Center "Al. Fleming" Institute of Immunology 14-16 Al.  
 Fleming street 16672 Vari, Greece ) whose assistance we gratefully





Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by David A. Hume (Dep't. of Biochemistry and Microbiology/Parasitology Institute for Molecular Biocscience University of Queensland Brisbane, Q 4072 Australia ) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

#### FEATURES

Location/Qualifiers  
1..351  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I830081B12"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_id="RIKEN full-length enriched, bone marrow macrophage"

#### BASE COUNT

72 a 93 c 113 g 73 t

Query Match 100.0%; Score 127; DB 13; Length 351;  
Best Local Similarity 100.0%; Pred. No. 1.8e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGCATCTGTGGCATCTGCAGGTCAGGTGATGATGCTGCTTCATGATCAAC 60  
DB 160 GTGGCATCTGTGGCATCTGCAGGTCAGGTGATGATGCTGCTTCATGATCAAC 219  
QY 61 TAAAAACAAGACAGAGACTGTGTGTGTGTGTGGGAGAGGTGTACCATTCCTTCACAA 120  
DB 220 TAAAAACAAGACAGAGACTGTGTGTGTGTGTGGGAGAGGTGTACCATTCCTTCACAA 279  
QY 121 CTGGCTGC 127  
DB 280 CTGGCTGC 286

RESULT 20 352 bp mRNA linear EST 07-DEC-2002  
LOCUS BY091208  
DEFINITION BY091208 RIKEN full-length enriched, 13 days embryo whole body Mus musculus cDNA clone K630085B17 5', mRNA sequence.  
ACCESSION BY091208

#### VERSION

BY091208.1 GI:26200367

#### KEYWORDS

EST.

Mus musculus (house mouse)

#### SOURCE

Mus musculus

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scleroglossa; Muridae; Mus.

#### REFERENCE

1 (bases 1 to 352)

#### AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamane, I., Kiyosawa, H., Tagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gojobori, T., Baladrelli, R., Hall, D.P., Bull, C., Hume, D.A., Quakebush, J., Schriml, L.M., Kanpin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Bruscic, V., Chochia, C., Cobani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Mik, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, C.A., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Borja, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasuniishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

#### FEATURES

Location/Qualifiers  
1..352





T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. 357

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="K630083012"

/tissue\_type="whole body"

/dev\_stage="13 days embryo"

/clone\_lib="RIKEN full-length enriched, 13 days embryo whole body"

## BASE COUNT

81 a 89 c 115 g 72 t

## ORIGIN

Query Match 100.0%; Score 127; DB 13; Length 357;  
Best Local Similarity 100.0%; Pred No. 1.8e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGATACCTGTGCATCTGCAGGTCAGATGATGATGCTGCTTCGATGTCAGC 60  
|||||  
Db 139 GTGCGATACCTGTGCATCTGCAGGTCAGATGATGATGCTGCTTCGATGTCAGC 198  
|||||

QY 61 TGAACAACAAGCAAGAGACTGTGTGTGTCTGGGAGAGTGAACCATTCCTCCACAA 120  
|||||  
Db 199 TGAACAACAAGCAAGAGACTGTGTGTGTCTGGGAGAGTGAACCATTCCTCCACAA 258  
|||||

QY 121 CTGCTGC 127  
|||||  
Db 259 CTGCTGC 265  
|||||

RESULT 23 BY302366 361 bp mRNA linear EST 11-DEC-2002  
LOCUS BY302366 RIKEN full-length enriched, 14.5 days embryo df/df  
DEFINITION Rathe's pouches Mus musculus cDNA clone K820003N18 5', mRNA  
sequence.  
ACCESSION BY302366  
VERSION BY302366  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 361)  
Okazaki, I., Osato, N., Saito, R., Suzuki, H., Yamataka, I., Kiyosawa, H., Nagai, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojohori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,

## TITLE

JOURNAL  
MEDLINE  
PUBMED

## COMMENT

Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Flecher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.D., Jarvis, E.D., Karai, A., Kawai, H., Kawasawa, Y., Kedierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pette, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sadelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verrado, R., Wagner, L., Walstedt, C., Wang, Y., Watanabe, Y., Wells, L., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carrington, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Haashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683  
12466851

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA ) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. 361

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="K820003N18"

/tissue\_type="Ratke's pouches"

/dev\_stage="14.5 days embryo df/df"

/clone\_lib="RIKEN full-length enriched, 14.5 days embryo df/df Ratke's pouches"

ORIGIN	BASE COUNT	71 a	99 c	118 g	73 t
Query Match	100.0%, Score 127, DB 13, Length 361,				
Best Local Similarity	100.0%, Pred. No. 1,8e-55,				
Matches 127, Conservative	0, Mismatches 0, Indels 0, Gaps 0				
QY	1	GTGGCATTCCTGGCATCTGGAGGTCACAGTGTATGCTGCTCCTTCGATGTCCAGC	60		
DB	177	GTGGCATTCCTGGCATCTGGAGGTCACAGTGTATGCTGCTCCTTCGATGTCCAGC	236		
QY	61	TGAATAACAAGCAAGAGACTGTGTGTGTGTGTGGGAGAGTGAACATTCTTCACAA	120		
DB	237	TGAATAACAAGCAAGAGACTGTGTGTGTGTGTGGGAGAGTGAACATTCTTCACAA	296		
QY	121	CTGCTGC 127			
DB	297	CTGCTGC 303			
RESULT 24					
AA403365					
LOCUS	AA403365	366 bp	mRNA	linear	EST 29-APR-1997
DEFINITION	va33a04.r1 Guaywoodford Beter mouse kidney day 7 Mus musculus cDNA clone IMAGE:733134 5' similar to WP:EK287.5 CB06614 ;, mRNA sequence.				
ACCESSION	AA403365				
VERSION	AA403365.1	GI:2057337			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus. 1 (bases 1 to 366)				
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE	The WashU-HHMI Mouse EST Project				
JOURNAL	Unpublished				
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINTL; contact the IMAGE Consortium (info@image.liml.gov) for further information. MG:450182				
FEATURES	High quality sequence stop: 317.				
source	Location/Qualifiers				
	1..366				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:733134"				
	/tissue_type="kidney"				
	/dev_stage="juvenile (7 days old)"				
	/lab_host="SOLR (kanamycin resistant)"				
	/clone_1ib="Guaywoodford Beter mouse kidney day 7"				
	/note="Organ: kidney; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3' Library provided by Lisa Guay-Woodford."				
BASE COUNT	87 a	84 c	114 g	81 t	
Query Match	100.0%, Score 127, DB 9, Length 366,				

	Best Local	Similarity	100.0%;	Pred.	No. 1.be-55;	
QY	Matches	127:	Conservative	0;	Mismatches	0; Indels 0; Gaps 0
Dd	1	GTCGATACCTGTGGCATCTTGCAAGGCTCAGGTGAATGATGCCTTCCATTGTCAAAGC	60			
QY	57	GTCGATACCCTGTGCCATCTGACAGGTCACAGTAGTAAGATCGCTTCGATGTCAAAGC	116			
Dd	117	TGAACAAGAGAAGAGCATCTGTGTGTGTCTGGAGAGAGTAAACCATTCCTTCACAAA	120			
QY	61	TGAACAAGAGAAGAGCATCTGTGTGTGTCTGGAGAGAGTAAACCATTCCTTCACAAA	120			
Dd	117	TGAACAAGAGAAGAGCATCTGTGTGTGTCTGGAGAGAGTAAACCATTCCTTCACAAA	120			
QY	121	CTGCTGC	127			
Dd	177	CTGCTGC	183			
RESULT 25						
LOCUS	AA144182					
DEFINITION	AA144182	367 bp	mRNA	linear	EST 18-FEB-1997	
ACCESSION	mg55bi0.xl Soares_thymus_2NBMT	Mus musculus cDNA clone IMAGE:582615				
VERSION	AA144182	5' similar to WP:ZK287.5 CE06614 ;,				
KEYWORDS	AA144182.1 GI:1713550					
SOURCE	EST.					
ORGANISM	Mus musculus (house mouse)					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubucq,T., Geisel,S., Kicaba,T., Lacy,W., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.					
TITLE	The WashU-HHMI Mouse EST Project					
JOURNAL	Unpublished					
COMMENT	Contact: Marras M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mousesest@watsn.wustl.edu This clone is available royalty-free through LNLT ; contact the IMAGE Consortium ( <a href="mailto:infoimage.lnl.gov">infoimage.lnl.gov</a> ) for further information. Seq primer: -28M13 rev2 from Amershams High quality sequence stop: 289. location/Qualifiers					
FEATURES	source					
	1..367					
	/organism="Mus musculus"					
	/mol_type="mRNA"					
	/strain="C57BL/6J"					
	/db_xref="taxon:10090"					
	/clone="IMAGE:582615"					
	/sex="male"					
	/tissue_type="Thymus"					
	/dev_stage="4 weeks"					
	/lab_host="DH10B"					
	/clone_id="Soares_thymus_2NBMT"					
	/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand CDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCACTAAGTGGAGAGCGCGCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."					
BASE COUNT	86 a	88 c	116 g	77 t		
ORIGIN						





Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)

1-1-22 Sueniyo-cho, Isurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Fax: 81-45-503-9216

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

URL: <http://genome.gsc.riken.go.jp/>

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
Location/Qualifiers  
1..376

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1830094D24"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_1ib="RIKEN full-length enriched, bone marrow  
macrophage"

BASE COUNT  
ORIGIN  
80 a 99 c 122 g 75 t

Query Match  
Best Local Similarity 100.0%; Score 127; DB 13; Length 376;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGATACCTGTCATCTGCAGGCTCCAGGTGATGATGCTGCTTCATGATCAAGC 60  
DB 151 GTGGATACCTGTCATCTGCAGGCTCCAGGTGATGATGCTGCTTCATGATCAAGC 210  
QY 61 TGAACAAGCAG 120  
DB 211 TGAACAAGCAG 270  
QY 121 CTGCTGC 127  
DB 271 CTGCTGC 277

RESULT 30  
LOCUS  
BY161786 382 bp mRNA linear EST 10-DEC-2002  
DEFINITION  
musculus cDNA clone 1830035A17 5', mRNA sequence.  
ACCESSION  
BY161786  
VERSION  
BY161786.1 GI:26298432  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 382)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bull, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Flecher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Secu, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation

JOURNAL  
MEDLINE  
PUBMED  
12466851  
COMMENT

of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

Alizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane  
, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Maki, K., Wataniki, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Submission

Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome, 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by David A. Hume (Dep't. of Biochemistry  
and Microbiology/Parasitology Institute for Molecular Bioscience  
University of Queensland Brisbane, Q 4072 Australia) whose  
assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source

Location/Qualifiers  
1..382  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1830035A17"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_1ib="RIKEN full-length enriched, bone marrow  
macrophage"

BASE COUNT  
ORIGIN  
80 a 98 c 126 g 78 t

Query Match  
Best Local Similarity 100.0%; Score 127; DB 13; Length 382;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGATACCTGTCATCTGCAGGCTCCAGGTGATGATGCTGCTTCATGATCAAGC 60  
DB 150 GTGGATACCTGTCATCTGCAGGCTCCAGGTGATGATGCTGCTTCATGATCAAGC 209  
QY 61 TGAACAAGCAG 120  
DB 210 TGAACAAGCAG 269  
QY 121 CTGCTGC 127  
DB 270 CTGCTGC 276

RESULT 31

BY307739 392 bp mRNA linear EST 11-DEC-2002  
 LOCUS BY307739 RIKEN full-length enriched, stroma cell Mus musculus cDNA  
 clone 1320001102 5', mRNA sequence.  
 ACCESSION BY307739  
 VERSION BY307739.1 GI:26498076  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 392)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,  
 Quackenbush, J., Schriml, L. W., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,  
 L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,  
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,  
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,  
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,  
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Piliat, R., Pontius, J. U., Qi, D.,  
 Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
 R. D., Tomita, M., Verrado, R., Wagner, L., Wahlestedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L. G., Wyszewski, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL NATURE  
 MEDLINE 22354683  
 PUBMED 12466851  
 COMMENT Contact: Yoshitake Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.c.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hirozane,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to

Prepare mouse tissues.  
 Tissues were provided by Takashi Ishikawa (Department of Surgery  
 2 Yokohama City University 3-9 Fukuura, Kanagawa-ku, Yokohama  
 236-0004 Japan) whose assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 Location/Qualifiers  
 source 1..392  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1320001102"  
 /cell\_type="stroma cell"  
 /clone\_1ib="RIKEN full-length enriched, stroma cell"  
 80 a 103 c 129 g 80 t  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 127; DB 13; Length 392;  
 Best local similarity 100.0%; Pred. No. 1.8e-55;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGCATACCTGTGCATCTGCAGGCTCAGGTGATGATGCTGCTTCATGTCACG 60  
 |||||||  
 DB 160 GTGCATACCTGTGCATCTGCAGGCTCAGGTGATGATGCTGCTTCATGTCACG 219  
 |||||||  
 QY 61 TGAACACAGCAGAGGACTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 120  
 |||||||  
 DB 220 TGAACACAGCAGAGGACTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 279  
 |||||||  
 QY 121 CTGCTGC 127  
 |||||||  
 DB 280 CTGCTGC 286  
 |||||||  
 RESULT 32  
 LOCUS CD565301 393 bp mRNA linear EST 11-JUN-2003  
 DEFINITION B0492H01-5 N1A Mouse E6.5 whole embryo cDNA library (long) Mus  
 musculus cDNA clone N1A:B0492H01 IMAGE:30447540 5', mRNA sequence.  
 ACCESSION CD565301  
 VERSION CD565301.1 GI:31609194  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 393)  
 Piao, Y., Ko, N. T., Lim, M. K. and Ko, M. S. H.  
 Construction of long-transcript enriched cDNA libraries from  
 submicrogram amounts of total RNAs by a universal PCR amplification  
 method  
 Genome Res. 11 (9), 1553-1558 (2001)  
 JOURNAL GENOME RESEARCH  
 MEDLINE 21429098  
 PUBMED 11544199  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: [cdna@gsun.grc.nia.nih.gov](mailto:cdna@gsun.grc.nia.nih.gov)  
 Plate: B0492 row: H column: 01  
 Seq primer: M13 Reverse  
 High quality sequence scop: 393  
 POLYA=No.  
 Location/Qualifiers  
 source 1..393  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="N1A:B0492H01 IMAGE:30447540"  
 /clone\_1ib="B0492H01 IMAGE:30447540"  
 /tissue\_type="E6.5 whole embryo"



```

/dev stage="whole embryo including extraembryonic tissues
at 6.5-days postcoitum"
/lab host="DH10B"
/clone lib="NIA Mouse E6.5 Whole Embryo CDNA Library (long
)
"/note="Vector: PCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse CDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://jgun.gsc.nia.nih.gov/cdna).
This is a long-transcript enriched CDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 1154191]). Total
RNAs were extracted from a pool of 7 embryos at 6.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
Oligo(dT) primer [Invitrogen:
5'-GCATGCTTCTAGATCGCAGCGCCCTTTTCTTTTCTTTT-3']
from 0.53 ug of total RNA, treated with T4 DNA polymerase,
and purified by ethanol-precipitation. The cDNAs were
ligated to lone-linker LP-SalI, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer SalI-4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of PCMV-SPORT6 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
2.3kb. The library was constructed by Yulan Plao."
BASE COUNT
ORIGIN
93 a 86 c 115 g 99 t
Query Match 100.0%; Score 127; DB 14; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.8e-55;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GTGGATACCTGTGCGCATCTGCAGGGTCAGGTGATGATGCGCTTCCTTCGATGCAAGC 60
Db 23 GTGCGATACCTGTGCGCATCTGCAGGGTCAGGTGATGATGCGCTTCCTTCGATGCAAGC 82
Oy 61 TGAAGAACAGCAAGAGGACTGTGTGTGTGTGTGGGAGAGTAAACATCTCTTCCACAA 120
Db 83 TGAAGAACAGCAAGAGGACTGTGTGTGTGTGTGGGAGAGTAAACATCTCTTCCACAA 142
Oy 121 CTGCTGC 127
Db 143 CTGCTGC 149
RESULT 33 394 bp mRNA linear EST 11-DEC-2002
BY278099
LOCUS BY278099 RIKEN full-length enriched, visual cortex Mus musculus
DEFINITION BY278099 RIKEN full-length enriched, visual cortex Mus musculus
ACCESSION BY278099
VERSION BY278099.1 GI:26468436
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
1 (bases 1 to 394)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oosato, N., Saito, R., Suzuki, H., Yamana, T., Kitayama, H.,
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
Quackenbush, J., Schreml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
Beisel, K.W., Blake, J.A., Brad, D., Bruscia, V., Chochia, C., Corbani,
L.E., Cousins, S., Dalla, E., Dragani, T.A., Flecher, C.F., Forrest,
A., Frazer, K.S., Gasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,
King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Mikl

```

```

H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perce, G.,
Pescio, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,
Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, U., Ring,
B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou,
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
R.D., Tomita, M., Verdaro, R., Wagner, L., Wahlstedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yangisawa,
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Waki, J., Aizawa, K.,
Aizawa, K., Aizawa, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851
COMMENT
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Saito-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
M., Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct
Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN Integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers
1..394
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430330L01"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT 81 a 99 c 130 g 82 t 2 others
ORIGIN
Query Match 100.0%; Score 127; DB 13; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.8e-55;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GTGGATACCTGTGCGCATCTGCAGGGTCAGGTGATGATGCGCTTCCTTCGATGCAAGC 60

```



Db 155 GTCGATACCTGTGTCATCTGCAGGTCAGGTGATGATGATGTCCTTCGATGTCACG 214  
 Qy 61 TGAACAACAGCAGAGAGACTGTGTGTGTGGAGAGAGTAAACATTCTTCACAA 120  
 Db 215 TGAACAACAGCAGAGAGACTGTGTGTGTGGAGAGAGTAAACATTCTTCACAA 274  
 Qy 121 CTGCTGC 127  
 Db 275 CTGCTGC 261  
 RESULT 34 395 bp mRNA 1linear EST 07-DEC-2002  
 BY095585  
 LOCUS BY095585  
 DEFINITION Musculus cDNA clone K630106W08 5', mRNA sequence.  
 ACCESSION BY095585  
 VERSION BY095585.1 GI:26206202  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 395)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, T., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schorbach, C., Gojohori, T., Baldarelli, R., Hill, D. P., Bulic, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, I., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbett, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. D., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shmada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyszynski, B. J., Yamaguchi, M., Yang, I., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Akawa, K., Arikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 TITLE  
 JOURNAL  
 MEDLINE 22354683  
 PUBMED 12466851  
 COMMENT  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsgami, M., Waki, K., Watanabe, K., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..395  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="K630106W08"  
 /tissue\_type="whole body"  
 /dev\_stage="17 days embryo"  
 /clone\_lib="RIKEN full-length enriched, 17 days embryo whole body"  
 BASE COUNT 88 a 97 c 128 g 81 t 1 others  
 ORIGIN  
 Query Match 100.0%; Score 127; DB 13; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-55;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GTCGATACCTGTGTCATCTGCAGGTCAGGTGATGATGATGTCCTTCGATGTCACG 60  
 Db 138 GTCGATACCTGTGTCATCTGCAGGTCAGGTGATGATGATGTCCTTCGATGTCACG 197  
 Qy 61 TGAACAACAGCAGAGAGACTGTGTGTGTGGAGAGAGTAAACATTCTTCACAA 120  
 Db 198 TGAACAACAGCAGAGAGACTGTGTGTGTGGAGAGAGTAAACATTCTTCACAA 257  
 Qy 121 CTGCTGC 127  
 Db 258 CTGCTGC 264  
 RESULT 35 397 bp mRNA 1linear EST 21-FEB-2001  
 BG277656  
 LOCUS ux47c07.y1 Soares\_NMAX\_maxillary process Mus musculus cDNA clone IMAGE:3513229 5', similar to TR:Q9WT21 Q9WT21 ZINC RING FINGER PROTEIN Smg. i, mRNA sequence.  
 DEFINITION  
 ACCESSION BG277656  
 VERSION BG277656.1 GI:13073177  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 397)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished  
 Other ESTs: ux47c07.x1  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:1393805  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 366.  
 FEATURES  
 Location/Qualifiers

source

1. 397

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:3513229"

/tissue\_type="maxillary process"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="Soares NMMA maxillary process"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5']

tggtaccatctgaaaggagcgccgcatgatacattcttcacaca 120

ttttacacatctgaaaggagcgccgcatgatacattcttcacaca 120

(Pharmacia), digested with Not I and Eco RI adaptors

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 94 a 94 c 125 g 84 t

ORIGIN

Query Match 100.0%; Score 127; DB 10; Length 397;

Best Local Similarity 100.0%; Pred. No. 1.8e-55;

Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGATACCTGTGCATCTGCAGGGTCCAGGTGATGATGCTGCTTCGATGCAAGC 60

DB 82 GTGCATACCTGTGCATCTGCAGGGTCCAGGTGATGATGCTGCTTCGATGCAAGC 141

QY 61 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 120

DB 142 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 201

QY 121 CTGCTGC 127

DB 202 CTGCTGC 208

RESULT 36

LOCUS BF011540 401 bp mRNA linear EST 06-OCT-2000

DEFINITION us40e01.y1 Soares NMMA branchial arch Mus musculus cDNA clone

IMAGE:3169560 5' similar to TR:Q9WTZ1 Q9WTZ1 ZINC RING FINGER

PROTEIN SAG. ; mRNA sequence.

BF011540

ACCESSION BF011540.1 GI:10711815

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 401)

REFERENCE NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

TITLE Unpublished

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

[image.llnl.gov/image/html/tresources.shtml](http://image.llnl.gov/image/html/tresources.shtml)

MGI:1065020

Seq primer: -40RP from GIBCO

High quality sequence stop: 394.

FEATURES

source

1. 401

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:3169560"

/tissue\_type="branchial arches"

/dev\_stage="embryo, 10.5 dpc"

/lab\_host="DH10B (phage resistant)"

/clone\_lib="Soares NMMA branchial arch"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5']

tggtaccatctgaaaggagcgccgcatgatacattcttcacaca 120

ttttacacatctgaaaggagcgccgcatgatacattcttcacaca 120

(Pharmacia), digested with Not I and Eco RI adaptors

I and Eco RI sites of the modified pT73 vector. Library

constructed and normalized by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 90 a 100 c 130 g 81 t

ORIGIN

Query Match 100.0%; Score 127; DB 10; Length 401;

Best Local Similarity 100.0%; Pred. No. 1.8e-55;

Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGATACCTGTGCATCTGCAGGGTCCAGGTGATGATGCTGCTTCGATGCAAGC 60

DB 134 GTGCATACCTGTGCATCTGCAGGGTCCAGGTGATGATGCTGCTTCGATGCAAGC 193

QY 61 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 120

DB 194 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGTGGGAGAGTGAACATTCCTTCACAA 253

QY 121 CTGCTGC 127

DB 254 CTGCTGC 260

RESULT 37

LOCUS W34101 409 bp mRNA linear EST 11-SEP-1996

DEFINITION mb01c09.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:318928 5' similar to WP:F35G12.9 CE00978 ; mRNA sequence.

W34101

ACCESSION W34101.1 GI:1316146

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 409)

REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The Mashu-HMI Mouse EST Project

JOURNAL The Mashu-HMI Mouse EST Project

COMMENT Contact: Marra, M/Mouse EST Project

Washu-HMI Mouse EST Project

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:209544

Seq primer: ETPprimer

High quality sequence stop: 408.

FEATURES

source

1. 409

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:318928"

/dev\_stage="19.5 dpc total fetus"

FEATURES	source	1. 414	/organism="Mus musculus"	/mol_type="rRNA"	/strain="C57BL/6J"	/db_xref="taxon:10090"	/clone="IMAGE:876702"	/sex="male"
FEATURES	source	1. 414	/organism="Mus musculus"	/mol_type="rRNA"	/strain="C57BL/6J"	/db_xref="taxon:10090"	/clone="IMAGE:876702"	/sex="male"

DB	QY	Query Match	Best Local Similarity	Matches 127; Conservative	Score 127; DB 9; Length 414;	100.0%;	100.0%;	Pred. No. 1.8e-55;	Mismatches 0;	Indels 0;	Gaps 0;
Db	QY	1	GTGGGATACCTGTGGCCATCTGCAGGGTCCAGGTGATGAGCCCTTGAGATCAAGC	60							
Db	QY	135	GTGGGATACCTGTGGCCATCTGCAGGGTCCAGGTGATGAGCCCTTGAGATCAAGC	194							
Db	QY	61	TGAAGACAGCAAGAGACTGTGTGGTGTGGTCTGGGGAGAGTGAACCATTCCTTCACAA	120							
Db	QY	195	TGAAGACAGCAAGAGACTGTGTGGTGTGGTCTGGGGAGAGTGAACCATTCCTTCACAA	254							
Db	QY	121	CTGCTGC	127							
Db	QY	255	CTGCTGC	261							

```

RESULT 39
BF469845 LOCUS
DEFINITION BF469845 416 bp mRNA linear EST 04-DEC-2000
UI-M-BH3-atw-f-06-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION UI-M-BH3-atw-f-06-0-UI 5', mRNA sequence.
VERSION BF469845
KEYWORDS BF469845.1 GI:11539076
SOURCE EST.
ORGANISM Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 416)
Bonaldo, M.F., Lemmon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7150, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m8t@mail.nih.gov
cDNA library preparation: M.B. Soares lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
location/Qualifiers
1. 416
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
FEATURES
SOURCE

```



Oy		1	GCGCAGTACCTGTGCAATCTGCAGGGTCACAGTTA TGATGCCCTTGCCATTGCATGTCAGC	60
Dd		144	GTCGACTACCCTGTGCATCTGAAGGCTCACAGTAGTAATGCCTGCTCATGTGCAGC	203
Oy		61	TGAATAACAAGAAGA CTGTGTGTGTGCTGTGGAGAGATGTAACCATTCCTTCACAA	120
Dd		204	TGAATAACAAGAAGA C G A C T GTGTGTGTGCTGTGGAGAGATGTAACCATTCCTTCACAA	263
Oy		121	CTGCTGC	127
Dd		264	CTGCTGC	270
RESULT_41	B8333776			
LOCUS	DEFINITION	B8333776	427 bp mRNA linear EST J-19-2000	
KEYWORDS	VERSION	us277h08.y1 Soares NMEDA branchial arch Mus musculus cDNA clone IMAGE:3168351 5'	similar to WP:R10A10.2 CEI2670 ZINC FINGER PROTEIN ; , mRNA sequence.	
ACCESSION	BE333776			
SOURCE	GI:9207552			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
JOURNAL COMMENT	Contact: Robert Strauszberg, Ph.D. Email: cgapbs-re@mail.nih.gov CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo			
FEATURES	Source			
	MG1:1063811			
	Seq primer:-40RP from Gibco			
	High quality sequence stop: 426.			
	Location/Qualifiers			
	1..427			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:3168351"			
	/tissue_type="branchial arches"			
	/dev_stage="embryo, 10.5 dpn"			
	/lab_host="DH10B (phage resistant)"			
	/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site 1: NotI; Site 2: EcoRI; 1st strand CDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCCAGCATTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."			
	BASE COUNT	100 a	97 c	134 g
	ORIGIN	96 t		
	Query Match	100.0%; Score 127; DB 10; Length 427;		
	Best Local Similarity	100.0%; Pred. No. 1.8e-55;		
	Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1	GTGGATACCTGTGCAATCTGCAGGGTCACAGTTA TGATGCCCTTGCCATTGCATGTCAGC	60	

Db		87	GTGGCATACCTGTCGCATCTGCAGAGGTCACAGGTAGTGATGCCCTTGCATGTCAAGC	146
OY		61	TGAATAACAAGAAAGAAGACTGTGTGTGGCTTGGGGAGAGTTAACCATTTCTTCCACAA	120
Db		147	TGAAAACAAGCAAGAGACTGTGTTGTGGCTTGGGGAGAGTGAACCATTTCTTCCACAA	206
OY		121	CTGCTGC	127
Db		207	CTGCTGC	213
RESULT 42				
LOCUS	B692461			
DEFINITION				
VERSION	B692461	429 bp	mRNA	linear
KEYWORDS	uw47g02.v1 Soares mammary gland NMLMG Mus musculus cDNA clone			
SOURCE	IMAGE:3465170 5' similar to WP:R10A10.2 CEI2670 ZINC FINGER PROTEIN			
ORGANISM	;; mRNA sequence.			
REFERENCE	B692461			
AUTHORS	B692461.1 GI:10080085			
TITLE	Mus musculus (house mouse)			
JOURNAL	Mus musculus			
COMMENT	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCI-CCAP http://www.ncbi.nlm.nih.gov/ccap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished Contact: Robert Straubeberg, Ph.D. Email: cgaps-f@mail.nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:1384530 Seq primer: -40RP from Gibco. Location/Qualifiers 1..429 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:3465170" /sex="female (lactating)" /tissue_type="mammary gland" /lab_host="DH10B" /clone_lib="Soares mammary gland NMLMG" /note="Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."			
BASE COUNT	97 a	104 c	138 g	90 t
ORIGIN				
Query Match	100.0%;	Score 127;	DB 10;	Length 429;
Best Local Similarity	100.0%;	Pred. No. 1.8e-55;		
Matches 127;	Conservative	0;	Mismatches	0;
OY	1	GTCGATACCTGTCGCATCTGCAGAGGTCACAGGTAGTGATGCCCTTGCATGTCAAGC	60	
Db	126	GTGGCATACCTGTGTCGCATCTGCAGAGGTCACAGGTAGTGATGCCCTTGCATGTCAAGC	185	
OY	61	TGAATAACAAGCAAGAGACTGTGTGTGGCTTGGGGAGAGTGAACCATTTCTTCCACAA	120	
Db	186	TGAAAACAAGCAAGAGACTGTGTGTGGCTTGGGGAGAGTGAACCATTTCTTCCACAA	245	
OY	121	CTGCTGC	127	
Db	246	CTGCTGC	252	

RESULT 43  
AA815479 440 bp mRNA linear EST 13-FEB-1998  
LOCUS  
DEFINITION  
vp20a07.r1 Soares mammary\_gland\_NBMWG Mus musculus cDNA clone  
IMAGE:1069140 5' similar to TR:P91404 P91404 SIMILARITY TO A3HC4  
ZINC FINGER ;, mRNA sequence.  
ACCESSION  
AA815479  
VERSION  
AA815479.1 GI:2885075  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 440)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished  
TITLE  
JOURNAL  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:591500  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 398.

# FEATURES

source  
1..440  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1069140"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares mammary gland\_NBMWG"  
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo (dT)  
primer (5'  
TGTTCACATCTGAAGTGGAGCGCGCGCATGTGTTTTTTTTTTTTTTTTTT  
T 3'); double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

## BASE COUNT

Query Match 100.0%; Score 127; DB 9; Length 440;  
Best Local Similarity 100.0%; Pred. No. 1.9e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
1 GTGGATACCTGTCATCTGACAGGTCAGGTGATGATGCTGCTTCATGATCAAGC 60  
DB 114 GTGGATACCTGTCATCTGACAGGTCAGGTGATGATGCTGCTTCATGATCAAGC 173  
OR 61 TGAACAAGCAAGAGACTGTGTGTGTGTGGGAGAGTGAACCATTCCTTCCACAA 120  
DB 174 TGAACAAGCAAGAGACTGTGTGTGTGTGTGGGAGAGTGAACCATTCCTTCCACAA 233  
OR 121 CTGCTGC 127  
|||||

DB 234 CTGCTGC 240

RESULT 44  
AA239898 443 bp mRNA linear EST 03-MAR-1997  
LOCUS  
DEFINITION  
mx61h04.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692791 5',  
similar to WP:ZKX87.5 CE06614 ;, mRNA sequence.  
ACCESSION  
AA239898  
VERSION  
AA239898.1 GI:1863937  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 443)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

The WashU-HMI Mouse EST Project  
Unpublished  
TITLE  
JOURNAL  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:426351  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 403.

# FEATURES

source  
1..443  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:692791"  
/tissue\_type="Liver"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse NML"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo (dT) primer (5'  
TGTTCACATCTGAAGTGGAGCGCGCGCATGTGTTTTTTTTTTTTTTTTTT  
T 3'); double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT  
ORIGIN  
100 a 106 c 144 g 93 t

Query Match 100.0%; Score 127; DB 9; Length 443;  
Best Local Similarity 100.0%; Pred. No. 1.9e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
1 GTGGATACCTGTCATCTGACAGGTCAGGTGATGATGCTGCTTCATGATCAAGC 60  
DB 133 GTGGATACCTGTCATCTGACAGGTCAGGTGATGATGCTGCTTCATGATCAAGC 192  
OR 61 TGAACAAGCAAGAGACTGTGTGTGTGTGGGAGAGTGAACCATTCCTTCCACAA 120  
DB 193 TGAACAAGCAAGAGACTGTGTGTGTGTGTGGGAGAGTGAACCATTCCTTCCACAA 252  
OR 121 CTGCTGC 127  
DB 253 CTGCTGC 259  
|||||

RESULT 45  
W34374 445 bp mRNA 1linear EST 13-MAY-1996  
LOCUS m56606.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:352523 5' similar to WP:F35G12.9 CE00978 ; mRNA  
sequence.  
ACCESSION W34374 GI:1316302  
VERSION W34374.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 445)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HM1 Mouse EST Project  
Unpublished  
Contact: Marra M/Mouse EST Project  
WashU-HM1 Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:224323  
Putative full length read  
Seq primer: EPRprimer  
High quality sequence stop: 444.  
Location/Qualifiers  
1..445  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:352523"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NBME13.5 14.5"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5].  
TGTTCACATCTGAAGTGGAGCGCGCCGGAATTTTCTTTTCTTTT  
T 3'), on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ); double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT7T3 vector. Library went through one round of  
normalization, and was constructed by Benito Soares and  
M. Fatima Bonaldo."

BASE COUNT 104 a 93 c 129 g 119 t  
ORIGIN  
Query Match 100.0%; Score 127; DB 14; Length 445;  
Best Local Similarity 100.0%; Pred. No. 1.9e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGATACCTGTGCATCTGCAGGCTCCAGGTGATGATGCTGCTTCGATGCAAC 60  
|||||  
Db 28 GTGCGATACCTGTGCATCTGCAGGCTCCAGGTGATGATGCTGCTTCGATGCAAC 87  
|||||  
QY 61 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGGGGAAGTGAACCATTCCTTCACAA 120  
|||||  
Db 88 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGGGGAAGTGAACCATTCCTTCACAA 147  
|||||

QY 121 CTGCTGC 127  
|||||  
Db 148 CTGCTGC 154  
|||||

RESULT 46  
AA036356 447 bp mRNA 1linear EST 26-AUG-1996  
LOCUS m744607.r1 Soares mouse p3NMFI9.5 Mus musculus cDNA clone  
DEFINITION IMAGE:472357 5' similar to WP:F35G12.9 CE00978 ; mRNA sequence.  
ACCESSION AA036356 GI:1509484  
VERSION AA036356  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 447)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HM1 Mouse EST Project  
Unpublished  
Contact: Marra M/Mouse EST Project  
WashU-HM1 Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:283101  
Seq primer: -28M13 rev2 from Amerham  
High quality sequence stop: 446.  
Location/Qualifiers  
1..447  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:472357"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares mouse p3NMFI9.5"  
/note="Vector: pT7T3D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5].  
TGTTCACATCTGAAGTGGAGCGCGCCGCAATTTTCTTTTCTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Benito  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

BASE COUNT 101 a 109 c 144 g 93 t  
ORIGIN  
Query Match 100.0%; Score 127; DB 9; Length 447;  
Best Local Similarity 100.0%; Pred. No. 1.9e-55;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGATACCTGTGCATCTGCAGGCTCCAGGTGATGATGCTGCTTCGATGCAAC 60  
|||||  
Db 135 GTGCGATACCTGTGCATCTGCAGGCTCCAGGTGATGATGCTGCTTCGATGCAAC 194  
|||||  
QY 61 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGGGGAAGTGAACCATTCCTTCACAA 120  
|||||  
Db 195 TGAACAACAAGCAAGAGACTGTGTGTGTGTGTGGGGAAGTGAACCATTCCTTCACAA 254  
|||||  
QY 121 CTGCTGC 127







Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,  
T., Inokuni, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Fagioli and Takao K. Hensch (   
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan ) whose  
assistance we gratefully acknowledge. Please visit our web site  
(http://genome.gsc.riken.go.jp) for further details.

## FEATURES

## source

1. 458  
location/Qualifiers

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="K330010J10"

/tissue\_type="visual cortex"

/clone\_fib="RIKEN full-length enriched, visual cortex"

BASE COUNT 101 a. 114 c 149 g 94 t

## ORIGIN

Query Match 100.0%; Score 127; DB 13; Length 458;  
Best Local Similarity 100.0%; Pred. NO. 1.9e-55;

Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGCAATACCTGTGTCATCTGCAGGTCAGGTGATGATGCTGCTTCGATGTCAGC 60
    |||||
DB 150 GTGCAATACCTGTGTCATCTGCAGGTCAGGTGATGATGCTGCTTCGATGTCAGC 209
    |||||

QY 61 TGAATAACAAGACAAGAGACTGTGTGTGTGCTGTGGAGAGAGTAAACATTCCTCCACA 120
    |||||
DB 210 TGAATAACAAGACAAGAGACTGTGTGTGTGCTGTGGAGAGAGTAAACATTCCTCCACA 269
    |||||

QY 121 CTGCTGC 127
    |||||
DB 270 CTGCTGC 276
```

Search completed: November 7, 2003, 11:53:53  
Job time : 1303.79 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 11:59:37 / Search time 1900.92 Seconds  
(without alignments)  
9640.351 Million cell updates/sec

Title: US-09-509-779-3

Perfect score: 754

Sequence: 1 ATGGCCGACGTGAGACGCG.....CAATTAAGTGCAGTTTAA 754

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estc1:\*  
10: gb\_estc2:\*  
11: gb\_estc3:\*  
12: gb\_estc4:\*  
13: gb\_estc5:\*  
14: em\_estfun:\*  
15: em\_estom:\*  
16: em\_estom:\*  
17: em\_esthum:\*  
18: em\_estinv:\*  
19: em\_estpin:\*  
20: em\_estvrt:\*  
21: em\_estfun:\*  
22: em\_estmam:\*  
23: em\_estmus:\*  
24: em\_estpro:\*  
25: em\_estrod:\*  
26: em\_estphg:\*  
27: em\_estvrt1:\*  
28: gb\_estc1:\*  
29: gb\_estc2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	689	91.4	1201	9 AL547435
2	683.6	90.7	853	13 BQ876591
3	681	90.3	1201	13 BX324742
4	664.8	88.2	949	12 BG769106

Result No.	Score	Query Match length	ID	Description
5	661.6	87.7	797	12 B1832643
6	659.8	87.5	918	12 B1759082
7	649	86.1	822	12 B1668735
8	647.4	85.9	818	12 B1457840
9	644.4	85.5	805	10 BQ708518
10	643	85.3	856	12 B1601855
11	640.2	84.9	754	12 BQ766992
12	635.4	84.3	1016	12 BMS56683
13	634.4	84.1	706	12 BM704660
14	634.4	84.1	724	12 B1226556
15	632.6	83.9	922	13 B192094
16	624.6	82.8	855	10 BE747000
17	623.6	82.7	893	10 BQ37017
18	619.4	82.1	1201	13 BX387707
19	616.2	81.7	797	12 B1858307
20	612.8	81.3	983	13 BQ648608
21	609	80.8	856	10 BQ753323
22	608.8	80.7	767	12 B1601470
23	607.4	80.6	702	9 A1338342
24	606.6	80.5	847	12 BQ773490
25	606	80.4	748	10 BQ76768
26	603.4	80.0	782	14 CB990062
27	597.2	79.2	904	10 BQ742338
28	595.8	79.0	692	14 CA449255
29	594.2	78.8	716	12 BQ761806
30	592.8	78.6	840	10 BQ037022
31	589.6	78.2	827	12 B1828930
32	588.8	78.1	675	12 B1858784
33	587.2	77.9	947	10 BQ111792
34	582.4	77.2	629	13 BQ632594
35	582.4	77.2	745	10 BQ121625
36	581.8	77.2	734	10 BQ397308
37	573.2	76.3	864	12 B1757881
38	574.8	76.2	793	10 BQ716023
39	571.8	75.8	691	10 BQ714665
40	566.4	75.1	641	13 BQ632301
41	559.4	74.2	678	13 BQ601136
42	559.2	74.2	946	14 CD556579
43	558.8	74.1	622	14 CB217926
44	558.2	74.0	656	13 BQ730720
45	555.4	73.7	595	10 BQ033587

## ALIGNMENTS

RESULT 1  
AL547435  
LOCUS  
DEFINITION  
AL547435 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
Clone CS0D1011YM21 5-PRIME, mRNA sequence.  
ACCESSION  
AL547435  
VERSION  
AL547435.2 GI:31269266  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 15, 2001 this sequence version replaced gi:12881506.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2013.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1011YM21&cluster=2013.f. Contact :  
Feng Liang Email : fliang@infotech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0D1011AG110P1.  
Location/Qualifiers

## FEATURES

source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1011M21"  
/tissue="PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 461 a 180 c 310 g 211 t 39 others  
ORIGIN

Query Match 91.4%; Score 689; DB 9; Length 1201;  
Best Local Similarity 97.5%; Pred. No. 4e-103;  
Matches 742; Conservative 1; Mismatches 11; Indels 7; Gaps 4;

1 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 60  
81 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 140  
61 GGCTCCAAAGTCGGGAGGCGACAAAGATGTTCTCCCTCAAGAAAGTGAACGGGTGGCCATG 120  
141 GGCTCCAAAGTCGGGAGGCGACAAAGATGTTCTCCCTCAAGAAAGTGAACGGGTGGCCATG 200  
121 TGGAGCTGGGACGTGGAAGTGCATGCGCCCATCTGCAAGGGTCCAGGTGATGATGCC 180  
201 TGGAGCTGGGACGTGGAAGTGCATGCGCCCATCTGCAAGGGTCCAGGTGATGATGCC 260  
181 TGTCTTAAGTGTCAAGCTGAAACAAACAGAGACTGTGTGTGCTGGGAGAAATGT 240  
261 TGTCTTAAGTGTCAAGCTGAAACAAACAGAGACTGTGTGTGCTGGGAGAAATGT 320  
241 AATCATTCCTTCCAAACCTGCGCATGTCCTCGTGGGAGAAACAAACATGCGCCCT 300  
321 AATCATTCCTTCCAAACCTGCGCATGTCCTCGTGGGAGAAACAAACATGCGCCCT 380  
301 CTCTCCAGCAGAGACTGGGTGTGTCCAAGAAATGAGAGTGTGTAAGAGCTTC 360  
381 CTCTCCAGCAGAGACTGGGTGTGTCCAAGAAATGAGAGTGTGTAAGAGCTTC 440  
361 TTAAGCGAGTGTGTAAGAGCTGTGATGATTCAGTGCCTTCAAAAGGCTTGA 420  
441 TTAAGCGAGTGTGTAAGAGCTGTGATGATTCAGTGCCTTCAAAAGGCTTGA 500  
421 ACACACAGGGGATGAATTCCTCAATAGAGAGCGATGATGATGTC--TTTGGACTC 478  
501 ACACACAGGGGATGAATTCCTCAATAGAGAGCGATGATGATGTC--TTTGGACTC 560  
479 ATCAAAGCCCTTGTT--AGCATTTGTGATTTTATCTCAGAAATTCCTGATTAAGA 536  
561 ATCAAAGCCCTTGTT--AGCATTTGTGATTTTATCTCAGAAATTCCTGATTAAGA 620  
537 AGATAATTATTAAGAGTGTGCTTCTCACTCTGTGTGTGTGTCGACACAGCTTA 596  
621 AGATAATTATTAAGAGTGTGCTTCTCACTCTGTGTGTGTGTCGACACAGCTTA 680  
597 GAAAGCTATTAAGAGAGAGAGAGCTCCAAATGTAATGACC--TTAATTTAAACCATTC 655  
681 GAAAGCTATTAAGAGAGAGAGAGCTCCAAATGTAATGACC--TTAATTTAAACCATTC 740  
656 TATAACAAGGACGTGAAGAGCTTC--GAGACTTTTTCATGCTTAATGTTGATCAGT 713  
741 TATAACAAGGACGTGAAGAGCTTC--GAGACTTTTTCATGCTTAATGTTGATCAGT 800  
714 TAAAAAAGATGTTTACATTAACAATAAAGTCAAGTTTAA 754  
801 TAAAAAAGATGTTTACATTAACAATAAAGTCAAGTTTAA 841

RESULT 2  
LOCUS BO876591 853 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGNCOURT 8584244 lupski\_sympathetic\_trunk Homo sapiens cDNA clone IMAGE:6192819 5', mRNA sequence.

ACCESSION BO876591  
VERSION BO876591.1 GI:22268599  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL unpublished  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: cgs@bbs-rcmail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM13595 row: h column: 04  
High quality sequence stop: 608.

FEATURES  
source  
1..853  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6192819"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult" 16 yr"  
/lab\_host="DH10B"  
/clone\_id="lupski\_sympathetic trunk"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1: NotI; Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGGCGTCCG-3' and 5'-GACTAGTTCATGATCGGAGCGGCCCTT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

BASE COUNT 221 a 191 c 225 g 205 t 1 others  
ORIGIN

Query Match 90.7%; Score 683.6; DB 13; Length 853;  
Best Local Similarity 97.1%; Pred. No. 1.1e-181;  
Matches 739; Conservative 0; Mismatches 15; Indels 7; Gaps 4;

1 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 60  
51 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 110  
61 GGCTCCAAAGTCGGGAGGCGACAAAGATGTTCTCCCTCAAGAAAGTGAACGGGTGGCCATG 120  
111 GGCTCCAAAGTCGGGAGGCGACAAAGATGTTCTCCCTCAAGAAAGTGAACGGGTGGCCATG 170  
121 TGGAGCTGGGACGTGGAAGTGCATGCGCCCATCTGCAAGGGTCCAGGTGATGATGCC 180  
171 TGGAGCTGGGACGTGGAAGTGCATGCGCCCATCTGCAAGGGTCCAGGTGATGATGCC 230  
181 TGTCTTAAGTGTCAAGCTGAAACAAACAGAGAGCTGTGTGTGCTGGGAGAAATGT 240  
231 TGTCTTAAGTGTCAAGCTGAAACAAACAGAGAGCTGTGTGTGCTGGGAGAAATGT 290  
241 AATCATTCCTTCCAAACCTGCGCATGTCCTCGTGGGAGAAACAGAAATGCGTGCCT 300

```

Db      291 AATCATCTCTCCACACTGCTGATGCTCCCTGGGTGAAACGAACTCGTGCCT 350
QY      301 CTGCGACGAGAGACTGGGTGCTCCAAAGATGCGCAATGAGGTGTTAAGAGCTTC 360
Db      351 CTGCGACGAGAGACTGGGTGCTCCAAAGATGCGCAATGAGGTGTTAAGAGCTTC 410
QY      361 TTAGCGAGTTGTTACAGAGCCCTGATGATCTGTATCCAGTCCCTACAAAGGCTTGA 420
Db      411 TTAGCGAGTTGTTACAGAGCCCTGATGATCTGTATCCAGTCCCTACAAAGGCTTGA 470
QY      421 ACACATACAGGAGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 478
Db      471 ACACATACAGGAGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 530
QY      479 ATCAAGACCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATCTCTGTATTAAG 536
Db      531 ATCAAGACCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATCTCTGTATTAAG 590
QY      537 AGATAATTTTAAAGGTGCTCTCTACCTCTGTGTGTGTGCGACAGACTTA 596
Db      591 AGATAATTTTAAAGGTGCTCTCTACCTCTGTGTGTGTGCGACAGACTTA 650
QY      597 GAAGTGTATTAAGAAAGAAAGAGCTTCAATTAATGATCACC--TTATATTTACCAATTC 655
Db      651 GAAGTGTATTAAGAAAGAAAGAGCTTCAATTAATGATCACC--TTATATTTACCAATTC 710
QY      656 TATACAAACGAGAGTGAAGCACTTTC--GAGACTTTTGTGATGCTTATGATGAT 713
Db      711 TATACAAACGAGAGTGAAGCACTTTC--GAGACTTTTGTGATGCTTATGATGAT 770
QY      714 TAAAAAAGATGTTACAGTAAATTAAGTGCAGTTTAA 754
Db      771 TAAAAAAGATGTTACAGTAAATTAAGTGCAGTTTAA 811

```

```

RESULT 3
LOCUS   BX324742 1201 bp mRNA linear EST 01-MAY-2003
DEFINITION BX324742 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS00D1037P13 5-PRIME, mRNA sequence.
ACCESSION BX324742
VERSION 1 GI:30307404
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1201)
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2013.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0A1037CH07Q1
&cluster=2013.f. Contact: Feng Liang Email: fliang@lifetech.com
URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Glenborough sequence ID: CS0A1037CH07Q1.
FEATURES
source
location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1037P13"
/risue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was

```

digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 90.3%; Score 681; DB 13; Length 1201;  
Best Local Similarity 96.5%; Pred. No. 7, 2e-181;  
Matches 734; Conservative 4; Mismatches 16; Indels 7; Gaps 4;

```

BASF COUNT 347 a 182 c 300 g 264 t 108 others
ORIGIN
QY      1 ATGGCCGAGGTGGAACGAGAGAAACCTGCGCTTCTCACTCCGGAGCTCA 60
Db      81 ATGGCCGAGGTGGAACGAGAGAAACCTGCGCTTCTCACTCCGGAGCTCA 140
QY      61 GGCTCAACCTGGGAGGCGCAAGAGTCTCCCTCAAGATGGAACCGCGGCGCAT 120
Db      141 GGCTCAACCTGGGAGGCGCAAGAGTCTCCCTCAAGATGGAACCGCGGCGCAT 200
QY      121 TGGAGCTGGGACGTGAGTGCATACGTGCGCATCTGACGGTCCAGGTGATGATGCC 180
Db      201 TGGAGCTGGGACGTGAGTGCATACGTGCGCATCTGACGGTCCAGGTGATGATGCC 260
QY      181 TGTCTTAGATGTCAGCTTAAGAAACAAACAGAGACTGTGTGTGTGGGAGATGT 240
Db      261 TGTCTTAGATGTCAGCTTAAGAAACAAACAGAGACTGTGTGTGTGGGAGATGT 320
QY      241 AATCATCTCTCCACAACTGTCGATGTCCTCTGTGGGTAAACAGAAACATGCTCCCT 300
Db      321 AATCATCTCTCCACAACTGTCGATGTCCTCTGTGGGTAAACAGAAACATGCTCCCT 380
QY      301 CTCTGCGACAGAGACTGGGTGCTCCAAAGATCGCAATGAGAGTGTAGAGGCTTC 360
Db      381 CTCTGCGACAGAGACTGGGTGCTCCAAAGATCGCAATGAGAGTGTAGAGGCTTC 440
QY      361 TTAGCGAGTTGTTACAGAGCCCTGTGTGATCTTTAATCAGTCCCTTCAAAAGGCTTGA 420
Db      441 TTAGCGAGTTGTTACAGAGCCCTGTGTGATCTTTAATCAGTCCCTTCAAAAGGCTTGA 500
QY      421 ACACATACAGGAGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 478
Db      501 ACACATACAGGAGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 560
QY      479 ATCAAGACCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATCTCTGTATTAAG 536
Db      561 ATCAAGACCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATCTCTGTATTAAG 620
QY      537 AGATAATTTTAAAGGTGCTCTCTACCTCTGTGTGTGTGCGACAGACTTA 596
Db      621 AGATAATTTTAAAGGTGCTCTCTACCTCTGTGTGTGTGCGACAGACTTA 680
QY      597 GAAGTGTATTAAGAAAGAAAGAGCTTCAATTAATGATCACC--TTATATTTACCAATTC 655
Db      681 GAAGTGTATTAAGAAAGAAAGAGCTTCAATTAATGATCACC--TTATATTTACCAATTC 740
QY      656 TATACAAACGAGAGTGAAGCACTTTC--GAGACTTTTGTGATGCTTATGATGATCAGT 713
Db      741 TATACAAACGAGAGTGAAGCACTTTC--GAGACTTTTGTGATGCTTATGATGATCAGT 800
QY      714 TAAAAAAGATGTTACAGTAAATTAAGTGCAGTTTAA 754
Db      801 TAAAAAAGATGTTACAGTAAATTAAGTGCAGTTTAA 841

```

```

RESULT 4
LOCUS   BG769106 949 bp mRNA linear EST 15-MAY-2001
DEFINITION BG769106 602743362F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4873095 5',
RNA sequence.
ACCESSION BG769106
VERSION 1 GI:14079759
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE  
1 (bases 1 to 949)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS  
NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCMP/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L10M1750 row: k column: 16  
High quality sequence stop: 748.  
Location/Qualifiers

FEATURES  
source

1..949  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4873095"  
/tissue\_type="melanotic melanoma, high MDR (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 49"  
/note="Organ: skin; Vector: pOT7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dt priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 272 a 260 g 213 t  
ORIGIN

Query Match 88.2%; Score 664.8; DB 12; Length 949;  
Best Local Similarity 97.3%; Pred. No. 2.4e-176;  
Matches 730; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

13 GAAGACGAGAGAAACCTGCGCCCTGCGCTCTCACTCCGGAGCTCAGGCTCAAGTCG 72  
1 GAAGACGAGAGAAACCTGCGCCCTGCGCTCTCACTCCGGAGCTCAGGCTCAAGTCG 60  
73 GGAGCGGACAGATGTTCTCCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAC 132  
61 GGAGCGGACAGATGTTCTCCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAC 120  
133 GTGAGTGGATACGTGCGCCATCTGACGGGTCCAGGTGATGGATGCTTATGATGT 192  
121 GTGAGTGGATACGTGCGCCATCTGACGGGTCCAGGTGATGGATGCTTATGATGT 180  
193 CAAGCTGAAACCAACAGAGAGATGTTGTGTCTGGGAGATGTATCATCTTC 252  
181 CAAGCTGAAACCAACAGAGAGATGTTGTGTCTGGGAGATGTATCATCTTC 240  
253 CACAACCTGCTGATGCTCTGTGGGTGAACAGAACATCGTGGCCCTCTGCCAGAG 312  
241 CACAACCTGCTGATGCTCTGTGGGTGAACAGAACATCGTGGCCCTCTGCCAGAG 300  
313 GACTGGGTGTCGCAAGAGATGGCAATGAGAGTGAAGGCTTCTTACGGCAGTTG 372  
301 GACTGGGTGTCGCAAGAGATGGCAATGAGAGTGAAGGCTTCTTACGGCAGTTG 360  
373 TTCAGAGCCCTGTGATCTTGTATCCAGTGCCTCAAGAGGCTAGAACACTACAGGG 432  
361 TTCAGAGCCCTGTGATCTTGTATCCAGTGCCTCAAGAGGCTAGAACACTACAGGG 420  
433 ATGAATTTTCAATAGAGCCGATGATCTGTGTCT--TTTGAATATCAAGCCTTG 490  
421 ATGAATTTTCAATAGAGCCGATGATCTGTGTCTTTTGGGACTCATCAAGCCTTG 480

QY 491 GTT--AGCATTTGTCAGTTTATCTTCAGAAATTCCTGTGATTAAGAGATTAATT 548  
DB 481 GTTAGCAATTTTGTGAGTTTATCTTCAGAAATTCCTGTGATTAAGAGATTAATT 540  
QY 549 AAAGTGTCTCTCTCACTCTGTGTGTGTGCGGACACAGCTTGAAGTGTATTA 608  
DB 541 AAAGTGTCTCTCTCACTCTGTGTGTGTGCGGACACAGCTTGAAGTGTATTA 600  
QY 609 AA-AGGAAAGAGCTCCAAATTAATCACC-TTATTAATTAACCAATTTCTATACACAG 666  
DB 601 AACAGGAAAGAGCTCCAAATTAATCACC-TTATTAATTAACCAATTTCTATACACAG 660  
QY 667 CAGTGAAGCAGTTTC--GAGCTTTTTCAGTGTGATGTTGATGTTGATTAAGAGAT 724  
DB 661 CAGTGAAGCAGTTTCAGAGAACTTTTGCATGCTTATGTTGATGTTGATTAAGAGAT 720  
QY 725 GTTACAGTACAAATTAAGTCACTTTAA 754  
DB 721 GTTACAGTACAAATTAAGTCACTTTAA 750

RESULT 5  
BI832643  
LOCUS  
DEFINITION  
603082021P1 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5221325 5',  
mRNA sequence.  
ACCESSION  
BI832643  
VERSION  
BI832643.1 GI:15944193  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
NIH-MGC <http://mgi.nci.nih.gov/>.  
1 (bases 1 to 797).  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished

Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L10M1556 row: 1 column: 06  
High quality sequence stop: 795.  
Location/Qualifiers

FEATURES  
source

1..797  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5221325"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_120"  
/note="Organ: pooled pancreas and spleen; Vector:  
pCW-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of spleen and pancreas from 28 yo  
male. Library is oligo-dt primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 1.5 kb, insert size range 1-2.5 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 025. Note: this is a NIH\_MGC Library."

BASE COUNT 208 a 179 c 212 g 198 t  
ORIGIN

Query Match 87.7%; Score 661.6; DB 12; Length 797;  
Best Local Similarity 96.9%; Pred. No. 1.8e-175;  
Matches 740; Conservative 0; Mismatches 14; Indels 10; Gaps 6;  
1 ATGGCCGACGTGAAGACGAGAGAAACCTGCGCCCTGCGCTCTCACTCCGGAGCTCA 60



Db	612	GAAGTCGTATAAAAAAGAAAGAGCCCAATTTGATCACTTTATATATTACCATTTTC	6711
Oy	656	TATACAAACGAGCAGTGGAAAGCAAGTTTC - GAGACTTTTCGATGCTTATGTTGATCAGT	713
Db	672	TATACAAACGGCAGTGGAAAGCAAGTTTCAGAGAACTTTTTCAGCTTATGTTGATCAGT	7211
Oy	714	TAA-AAAAGATGTTACAGTTAACCAATTAAGTG	745
Db	732	TACCAAAAGATGTTACAGTTAACCAATTTAGAG	764

RESULT 7	LOCUS	DEFINITION
BI668735	BI668735	822 bp mRNA linear EST 12-SEP-2001
	603293174741	NIH_MGC_96 Homo sapiens CDNA clone IMAGE:5312226 5', mRNA sequence.

VERSION	B1668735.1	GI:15582968
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE	1 (bases 1 to 822)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NNGRI), Shikui  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.lnlnl.gov>  
 Plate: LAM11791 row: d column: 19  
 High quality sequence stop: 788.

FEATURES	Location/Qualifiers
source	1. .822

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:531222c"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/clone_idb="NIH_MGC_96"
/notice="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (cgcagc
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

```

[illegible]

QY	121	TGAGACTGGGACGTGGAGTGCATATCGTCGCATCTGGAGGGTCCAGGTGATGATGCC	180
Db	149	TGGAGCTGGGACGTGGAGTGCATATCGTCGCATCTGGAGGGTCCAGGTGATGATGCC	208
QY	181	TGCTTTAATGTCACAGCTGAAAAACAACAAGAGACTGTGTTGTGGTCTGGGGAGATGT	240
Db	209	TGCTTTAATGTCACAGCTGAAAAACAACAAGAGACTGTGTTGTGGTCTGGGGAGATGT	268
QY	241	AATCATTCCTTCACAACTGCTGCATGTCCTCTGGGGTAAAACAGAACATCGCTCCCT	300
Db	269	AATCATTCCTTCACAACTGCTGCATGTCCTCTGGGGTAAAACAGAACATCGCTCCCT	328
QY	301	CTGCGCCACAGGACTGGGTGGTCCAAAGAAATCGGCAATGAGAGTGTTAAGAGCTTC	360
Db	329	CTGCGCCACAGGACTGGGTGGTCCAAAGAAATCGGCAATGAGAGTGTTAAGAGCTTC	388
QY	361	TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATATCAGTGCCTTCAAAAGCTTAG	420
Db	389	TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATATCAGTGCCTTCAAAAGCTTAG	448
QY	421	ACACTACAGGGGATGAAATTTCTCAATATGAGACCGATGATCTGTGTC- -TTTGGACTC	478
Db	449	ACACTACAGTGAATTAATTTCTCAATATGAGACCGATGATCTGTGTCCTTTGGAGTC	508
QY	479	ATCAAAAGCCTTGTT- -AGCATTTGTCAAGTTTATCTTCAGAAATCTCTGTGATTAGA	536
Db	509	ATCAAAAGCCTTGTTTACATTTTGTCTAGTTTATCTTCAGAAATCTCTGTGATTAGA	568
QY	537	AGATTAATTTTAAAGTGTGTCCTTCACACCTCTGTGGTGTGTGCGGCACACAGCTTA	596
Db	569	AGATTAATTTTAAAGTGTGTCCTTCACACCTCTGTGGTGTGTGCGGCACACAGCTTA	628
QY	597	GAAGTGCTATAAAAA-AGGAAAGAGCTCCAAATTGAATCAC- -TTATAATTTTACCATT	654
Db	629	GAAGTGCTATAAAAAAGGAAAGAGCTCCAAATTGAATCACCTTATAATTTTACCATT	688
QY	655	CTATACACACAGGACGTGGAGACAGTTTCGAG- -ACTTTTTCGATGCTTATGTTGATCAGT	713
Db	689	CTATACACACAGGACGTGGAGACAGTTTCGAGACCTTTTTCGATGCTTATGTTGATCAGT	748
QY	714	TAAAAAAGATGCTATACGTAACAATAAATGTCAGTTTAAA	754
Db	749	T-ACCAAGAAATGTTACAGTAACCAAT- -AAGTGCAGTTTAAA	787

RESULT 8  
LOCUS B1457840  
DEFINITION B1457840 818 bp mRNA EST 21-AUG-200  
603199212.F1 NIH\_MGC\_96 Homo sapiens CDNA clone IMAGE:5277723 5',  
mRNA sequence.  
B1457840  
B1457840  
B1457840.1 GI:15248496  
EST.  
Homo sapiens (human)  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 818)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-romail.nih.gov](mailto:cgaps-romail.nih.gov)  
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11701 row: g column: 04  
High quality sequence stop: 80.



FEATURES  
Source

Location/Qualifiers  
1. .818  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5277723"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 225 a 174 c 214 g 205 t  
ORIGIN

Query Match 85.9%; Score 647.4; DB 12; Length 818;  
Best Local Similarity 95.6%; Pred. No. 1.9e-171;  
Matches 733; Conservative 0; Mismatches 21; Indels 13; Gaps 6;

1 ATGGCCGACGTGGAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 60  
17 ATGGCCGACGTGGAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 76  
61 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 120  
77 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 136  
121 TGGAGCTGGAGACGTGAGAGTCCGATAGTGGCCATCTGCGAGGCTCCAGGTGATGCC 180  
137 TGGAGCTGGAGACGTGAGAGTCCGATAGTGGCCATCTGCGAGGCTCCAGGTGATGCC 196  
181 TGTCTTAGATGTCAACCTGAAACAAACAAAGAGACTGTGTGTGGTCTGGGGAGATGT 240  
197 TGTCTTAGATGTCAACCTGAAACAAACAAAGAGACTGTGTGTGGTCTGGGGAGATGT 256  
241 AATCATTTCTTCCACCACTGCTGCATGTCTGTGGGTGAACAGAACATCGCTCCCT 300  
257 AATCATTTCTTCCACCACTGCTGCATGTCTGTGGGTGAACAGAACATCGCTCCCT 316  
301 CTCTGCGACAGACCTGGGTGTCCAAAGATGCGCAATGAGAGTGTGAAGAGCTTC 360  
317 CTCTGCGACAGACCTGGGTGTCCAAAGATGCGCAATGAGAGTGTGAAGAGCTTC 376  
361 TTAGCCGACGTGTTGAGAGCCCTGGTGAATCTGTAATCCAGTCCCTCAAGAGCTAGA 420  
377 TTAGCCGACGTGTTGAGAGCCCTGGTGAATCTGTAATCCAGTCCCTCAAGAGCTAGA 436  
421 AACTAGACAGGGAGTGAATTTCTTCAATAGAGCCGATGATCTGTGTC--TTTGGATC 478  
437 AACTAGACAGGGAGTGAATTTCTTCAATAGAGCCGATGATCTGTGTC--TTTGGATC 496  
479 ATCAAGACCTTGGTT--AGCATTTGTCAGTTTATCTTCAAGAAATCTCTGCATTTAAG 536  
497 ATCAAGACCTTGGTT--AGCATTTGTCAGTTTATCTTCAAGAAATCTCTGCATTTAAG 556  
537 AGATAATTTATTAAGGTGATCTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596  
557 AGATAATTTATTAAGGTGATCTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 616  
597 GAAGTCTATTAAGAAAGAAAGAGCTCAAAATGAATCAAC--TTATATTTTACCAATTC 655  
617 GAAGTCTATTAAGAAAGAAAGAGCTCAAAATGAATCAAC--TTATATTTTACCAATTC 676  
656 TATAC--AAAGAGAGTGAAGAGCTTC--GAGACTTTTTCAGTCTTATGTTGATTC 710  
677 TATACAAAGAGAGTGAAGAGCTTC--GAGACTTTTTCAGTCTTATGTTGATTC 736

QY 711 AGTT---AAAAAGATTTACAGTAAACAATAAGTGAAGTTAA 754  
Db 737 CAGTTAAAGAAAGATTTACAGTAAACAATAAGTGAAGTTAA 783

RESULT 9  
BG708518  
LOCUS  
DEFINITION  
602670411F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4793091 5', mRNA sequence.  
ACCESSION  
BG708518  
VERSION  
BG708518.1 GI:13985940  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLNL0672 row: f column: 04  
High quality sequence stop: 740.

FEATURES  
source

Location/Qualifiers  
1. .805  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4793091"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 214 a 183 c 214 g 194 t  
ORIGIN

Query Match 85.5%; Score 644.4; DB 10; Length 805;  
Best Local Similarity 94.6%; Pred. No. 1.3e-170;  
Matches 723; Conservative 0; Mismatches 31; Indels 10; Gaps 5;

1 ATGGCCGACGTGGAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 60  
26 ATGGCCGACGTGGAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 85  
61 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 120  
86 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 145  
121 TGGAGCTGGAGACGTGAGAGTCCGATAGTGGCCATCTGCGAGGCTCCAGGTGATGCC 180  
146 TGGAGCTGGAGACGTGAGAGTCCGATAGTGGCCATCTGCGAGGCTCCAGGTGATGCC 205  
181 TGTCTTAGATGTCAAGCTGAAACAAACAAAGAGACTGTGTGTGTGTGTGTGTGTGTGTGT 240  
206 TGTCTTAGATGTCAAGCTGAAACAAACAAAGAGACTGTGTGTGTGTGTGTGTGTGTGTGT 265

QY 241 AATCATCTCTTCCACACACTGCTGATGTCCCTGTGGTGAACAGAACATGCTGCTT 300  
 DB 266 AATCATCTCTTCCACACACTGCTGATGTCCCTGTGGTGAACAGAACATGCTGCTT 325  
 QY 301 CTCCTGACAGAGACTGGGTGGTCCAAAGATCGGCAATGAGAGGTTAGAGGCTTC 360  
 DB 326 CTCCTGACAGAGACTGGGTGGTCCAAAGATCGGCAATGAGAGGTTAGAGGCTTC 385  
 QY 361 TTAGGCGACTTGTTCAGAGCCCTGTGTGATCTTGTATCAGTGGCTTCAAAAGCTTGA 420  
 DB 386 TTAGGCGACTTGTTCAGAGCCCTGTGTGATCTTGTATCAGTGGCTTCAAAAGCTTGA 445  
 QY 421 ACACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 478  
 DB 446 ACACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 505  
 QY 479 ACACAAAGCTTGGT--AGCATTTGTCAATTTATCTTCAAGAAATCTCTGTGATTAAGA 536  
 DB 506 ATCAAAAGCTTGGT--AGCATTTGTCAATTTATCTTCAAGAAATCTCTGTGATTAAGA 565  
 QY 537 AGATATTTATTTAA--AGTGTCTCTTCTTACCTCTGTGTGTGTGTGCGCACACAGCTT 595  
 DB 566 AGATATTTATTTAA--AGTGTCTCTTCTTACCTCTGTGTGTGTGTGCGCACACAGCTT 625  
 QY 596 AGAAGTCTATTAATAAAGGAAAGAGTCCAAATGATATCAC--TTTATATTAACCATTT 654  
 DB 626 AGAAGTCTATTAATAAAGGAAAGAGTCCAAATGATATCAC--TTTATATTAACCATTT 685  
 QY 655 CTATACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 710  
 DB 686 CTATACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 745  
 QY 711 AGTTAAAAAAGATGTTACAGTACCAATTAAGTGCATTTAA 754  
 DB 746 AGTTAAAAAAGATGTTACAGTACCAATTAAGTGCATTTAA 789

RESULT 10  
 B1601855 856 bp mRNA linear EST 07-SHP-2001  
 LOCUS 603244729F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5287014 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1601855  
 VERSION B1601855.1 GI:15494794  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 856)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsabrs@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshitaki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM11725 row: j column: 07  
 High quality sequence step: 820.  
 Location/Qualifiers  
 1. 856  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5287014"  
 /tissue\_type="hypothalamus"

/lab host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /note="Organ: Brain; Vector: Bluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtagag  
 ); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTTV-3',  
 size selected for average insert size 2.3 kb and  
 normalized to 10^5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 227 a 186 c 227 g 216 t  
 ORIGIN

Query Match 85.3%; Score 643; DB 12; Length 856;  
 Best Local Similarity 95.3%; Pred. No. 3.4e-170;  
 Matches 729; Conservative 0; Mismatches 25; Indels 11; Gaps 6;

QY 1 ATGGCCGACGTGGAAAGACGAGAGAGAAACCTGGCCCTGCTCACTCCGGAGCTCA 60  
 DB 17 ATGGCCGACGTGGAAAGACGAGAGAGAAACCTGGCCCTGCTCACTCCGGAGCTCA 76  
 QY 61 GGCCTCAAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGAGTGAACGCGGTGGCCATG 120  
 DB 77 GGCCTCAAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGAGTGAACGCGGTGGCCATG 136  
 QY 121 TGAAGCTGGGACGTGAGTGCATACGTGCCCATCTGACAGGCTCAGGTATGATGCC 180  
 DB 137 TGAAGCTGGGACGTGAGTGCATACGTGCCCATCTGACAGGCTCAGGTATGATGCC 196  
 QY 181 TGTCTTAGATGTCAAGCTGAAACCAACCAAGAGAGCTGTGTGTGTGGGAGAAATGT 240  
 DB 197 TGTCTTAGATGTCAAGCTGAAACCAACCAAGAGAGCTGTGTGTGTGGGAGAAATGT 256  
 QY 241 AATCATCTCTTCCACACACTGCTGATGTCCCTGTGGTGAACAGAACATGCTGCTT 300  
 DB 257 AATCATCTCTTCCACACACTGCTGATGTCCCTGTGGTGAACAGAACATGCTGCTT 316  
 QY 301 CTCCTGACAGAGACTGGGTGGTCCAAAGATCGGCAATGAGAGGTTAGAGGCTTC 360  
 DB 317 CTCCTGACAGAGACTGGGTGGTCCAAAGATCGGCAATGAGAGGTTAGAGGCTTC 376  
 QY 361 TTAGGCGACTTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGGCTTCAAAAGCTTGA 420  
 DB 377 TTAGGCGACTTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGGCTTCAAAAGCTTGA 436  
 QY 421 ACACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 477  
 DB 437 ACACACAGAGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 496  
 QY 478 CATCAAGCTTGGT--AGCATTTGTCAATTTATCTTCAAGAAATCTCTGTGATTAAG 535  
 DB 497 CATCAAGCTTGGT--AGCATTTGTCAATTTATCTTCAAGAAATCTCTGTGATTAAG 556  
 QY 536 AAGATATTTATTTAA--AGTGTCTCTTCTTACCTCTGTGTGTGTGCGCACACAGCT 594  
 DB 557 AAGATATTTATTTAA--AGTGTCTCTTCTTACCTCTGTGTGTGTGCGCACACAGCT 616  
 QY 595 TAGAAGTCTATTAATAAAGGAAAGAGCTCCAAATGAAATCAC--TTTATATTTAACCATT 653  
 DB 617 TAGAAGTCTATTAATAAAGGAAAGAGCTCCAAATGAAATCAC--TTTATATTTAACCATT 676  
 QY 654 TCTATACACAGGACGTGGAGAGCTTTC--GAGACTTTTTCATGCTTATAG--TTGAT 709  
 DB 677 TCTATACACAGGACGTGGAGAGCTTTC--GAGACTTTTTCATGCTTATAG--TTGAT 736  
 QY 710 CAGTTAAAAAAGATGTTACAGTAACTTAAAGTGCATTTAA 754  
 DB 737 CAGTTAAAAAAGATGTTACAGTAACTTAAAGTGCATTTAA 781

RESULT 11  
 BG766992

LOCUS BG766992 754 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602740511F1 NIH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4870251 5',  
 mRNA sequence.  
 ACCESSION BG766992  
 VERSION BG766992.1 GI:14077645  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 754)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC/DCTD/DRP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.jnl.gov>  
 Plate: LLCM1743 row: e column: 04  
 High quality sequence stop: 750.  
 Location/Qualifiers  
 1..754  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4870251"  
 /tissue\_type="melanotic melanoma, high MDR (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_49"  
 /note="Organ: Skin; Vector: pOT7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 205 a 160 c 197 g 192 t

Query Match 84.9%; Score 640.2; DB 12; Length 754;  
 Best Local Similarity 96.9%; Pred. No. 2e-163;  
 Matches 728; Conservative 0; Mismatches 13; Indels 10; Gaps 7;

13 GAAGAGCGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCAGGCTTCAAGTCG 72  
 |||||  
 1 GAAGAGCGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCAGGCTTCAAGTCG 60  
 |||||  
 73 GGAGCGCAAGAATGTTCTCCCTCAAGAAGTGAACGGGTGGCCATGTGGAGCTGGAC 132  
 |||||  
 61 GGAGCGCAAGAATGTTCTCCCTCAAGAAGTGAACGGGTGGCCATGTGGAGCTGGAC 120  
 |||||  
 133 GTGAGTGGCATCTGCGGCATCTGCAGGGTCCAGGTATGATGATCTTTAGATGT 192  
 |||||  
 121 GTGAGTGGCATCTGCGGCATCTGCAGGGTCCAGGTATGATGATCTTTAGATGT 180  
 |||||  
 193 CAAGCTGAAAACAAACAAAGAGAGCTGTTGTGTGTGGGAGATGTATCATCTTC 252  
 |||||  
 181 CAAGCTGAAAACAAACAAAGAGAGCTGTTGTGTGTGGGAGATGTATCATCTTC 240  
 |||||  
 253 CCAACTGCTGCATGTCCTGTGGTGAACAGAAATGCGTCCCTCTCCGACGAG 312  
 |||||  
 241 CCAACTGCTGCATGTCCTGTGGTGAACAGAAATGCGTCCCTCTCCGACGAG 300  
 |||||  
 313 GACTGGTGTCTCCAAAGATCGGCAATGAGAGTGTAGAGGCTTTAGGCGAGTTG 372  
 |||||  
 301 GACTGGTGTCTCCAAAGATCGGCAATGAGAGTGTAGAGGCTTTAGGCGAGTTG 360  
 |||||

373 TTCAGAGCCCTGGTGGATCTTGTATTCAGTGCCTTCAAGGCTAGAACACTACAGGGG 432  
 |||||  
 361 TTCAGAGCCCTGGTGGATCTTGTATTCAGTGCCTTCAAGGCTAGAACACTACAGGGG 420  
 |||||  
 433 ATGAATCTTCAAAATAGAGCCGATGATCTGTGTC--TTTGACTCATCAAGCCTTG 490  
 |||||  
 422 ATGAATCTTCAAAATAGAGCCGATGATCTGTGTC--TTTGACTCATCAAGCCTTG 480  
 |||||  
 491 GTT--AGCATTTGTCACTTTATCTTCAAGAAATCTCTGTGATTA--GAAGATATTTAT 547  
 |||||  
 481 GTTACCATTTTGTCACTTTTATCTTCAAGAAATCTCTGTGATTA--GAAGATATTTAT 540  
 |||||  
 548 TAA--GGTGTCTCTTCTACCTCTGTGGTGTGTGGCCACACAGCTTGAAGTCTAT 606  
 |||||  
 541 TAAAGGTGTCTCTTCTACCTCTGTGGTGTGTGGCCACACAGCTTGAAGTCTAT 600  
 |||||  
 607 AAAAAAGAAAGAGCTCCAAATTTGAATCAAC--TTATATTTACCCATTTCTATCAACAG 665  
 |||||  
 601 AAAAAAGAAAGAGCTCCAAATTTGAATCAAC--TTATATTTACCCATTTCTATCAACAG 660  
 |||||  
 666 GCAGTGGAAGCAGTTTC--GAACCTTTTCAGTCTTATGTTGATCAGTTAAAAAGAA 723  
 |||||  
 661 GCAGTGGAAGCAGTCTCAGAGAACTTTTGATGCTTATGTTGATCAGTTAAACAAGAA 720  
 |||||  
 724 TGTACAGTAACAATAAAGTCACTTTAA 754  
 |||||  
 721 TGTACAGTAACAATAAAGTCACTTTAA 750  
 |||||

RESULT 12  
 BMS56683 1016 bp mRNA linear EST 20-FEB-2002  
 LOCUS BMS56683  
 DEFINITION AGENCOURT\_6540688 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5737890  
 5', mRNA sequence.  
 ACCESSION BMS56683  
 VERSION BMS56683.1 GI:18798111  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1016)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.jnl.gov>  
 Plate: LLAH12748 row: 1 column: 19  
 High quality sequence stop: 620.  
 Location/Qualifiers  
 1..1016  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5737890"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_88"  
 /note="Organ: small intestine; Vector: pCMV-SPORT6;  
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
 oligo-dT primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 271 a 231 c 266 g 246 t

Query Match 84.3%; Score 635.4; DB 12; Length 1016;  
 Best Local Similarity 95.5%; Pred. No. 5e-168;  
 Matches 721; Conservative 0; Mismatches 21; Indels 13; Gaps 6;

1 ATGGCCGACGCTGGAGAGAGAGAACTCGCCCTGCGCTTCACATCCGGAGACTCA 60  
 |||||  
 52 ATGGCCGACGCTGGAGAGAGAGAACTCGCCCTGCGCTTCACATCCGGAGACTCA 111  
 |||||  
 61 GGCTCCAGATCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 120  
 |||||  
 112 GGCTCCAGATCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 171  
 |||||  
 121 TGGAGCTGGAGAGTGGAGTGCATATCGTCCCATCTGAGGGTCCAGTGAATGATGCC 180  
 |||||  
 172 TGGAGCTGGAGAGTGGAGTGCATATCGTCCCATCTGAGGGTCCAGTGAATGATGCC 231  
 |||||  
 181 TGTCTAGATGTCAGAGCTGAAAACAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 240  
 |||||  
 222 TGTCTAGATGTCAGAGCTGAAAACAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 291  
 |||||  
 241 AATCATTCCTTCACAACTGCTGATGTCCTGCTGGGTGAACAGAACATCGTGCCT 300  
 |||||  
 292 AATCATTCCTTCACAACTGCTGATGTCCTGCTGGGTGAACAGAACATCGTGCCT 351  
 |||||  
 301 CTCTGCGACGAGAGTGGTGGTCCAAAGATCGGCAATGAGAGTGTGTAAGGCTTC 360  
 |||||  
 352 CTCTGCGACGAGAGTGGTGGTCCAAAGATCGGCAATGAGAGTGTGTAAGGCTTC 411  
 |||||  
 361 TTAGGCGAGTGTTCAGAGCCCTGGTGGATCTGTATATCAGTGGCCCTCAAGGCTGA 420  
 |||||  
 412 TTAGGCGAGTGTTCAGAGCCCTGGTGGATCTGTATATCAGTGGCCCTCAAGGCTGA 471  
 |||||  
 421 ACACACTACAGGGGATGATTTCTTCAATAGAGCCGATGATGCTGTGTC- -TTTGGACTC 478  
 |||||  
 472 ACACACTACAGGGGATGATTTCTTCAATAGAGCCGATGATGCTGTGTC- -TTTGGACTC 531  
 |||||  
 479 ATCAAAAGCTTGGT- -AGCATTTGTCAAGTTTATCTTCAAGAAATTTCTGTGATTAAGA 536  
 |||||  
 532 ATCAAAAGCTTGGTGGTGGATTTGTTCAGTTTATCTTCAAGAAATTTCTGTGATTAAGA 591  
 |||||  
 537 AGATATATTATTAAGGTGGTCTTCTCAACCTCTGT- -GGTGTGTGGGCGACAGAGCT 595  
 |||||  
 552 AGATATATTATTAAGGTGGTCTTCTCAACCTCTGTGTGTGTGTGTGTGTGTGTGTGT 651  
 |||||  
 556 AGAAGTGTATTAAGGAGAGAGAGAGCTCCAAATTGATAC- -TATATATTATTTACCATTT 654  
 |||||  
 652 AGAAGTGTATTAAGGAGAGAGAGAGAGCTCCAAATTGATACCTTTATATTTACCATTT 711  
 |||||  
 655 CTATACCAACAGGAGTGAAGAGAGTTTGA- -CTTTTGTGATGCTTATGTTGA 708  
 |||||  
 712 CTATACCAACAGGAGTGAAGAGAGTTTGAAGAACTTTTTCATGAGGCTTATGTTGA 771  
 |||||  
 709 TCAG- -TTAAAAAGAGATGTTACATTAACAATAA 742  
 |||||  
 772 TCAGTTTAAAAAGAGATGTTACATTAACAATAA 806  
 |||||

RESULT 13  
 BM704660 706 bp mRNA linear EST 28-FEB-2002  
 LOCUS UT-E-C11-af0-o-15-0-UI.r2 UT-E-C11 Homo sapiens cDNA clone  
 DEFINITION UT-E-C11-af0-o-15-0-UI 5', mRNA sequence.  
 ACCESSION BM704660  
 VERSION BM704660.1 GI:19017918  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 706)  
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery  
 MEDLINE Genome Res. 6 (9), 791-806 (1996)  
 PUBMED 97044477  
 COMMENT 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@iowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..706  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UT-E-C11-af0-o-15-0-UI"  
 /tissue\_type="RPE and Choroid"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UT-E-C11"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UT-E-C11 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Ronaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 182 a 153 c 191 g 180 t

ORIGIN

Query Match 84.1%; Score 634.4; DB 12; Length 706;  
 Best Local Similarity 97.5%; Pred. No. 8.3e-168;  
 Matches 688; Conservative 0; Mismatches 11; Indels 7; Gaps 4;

5 CCGACGTGGAAGAGAGAGAGAACTGCGCCCTGCTCATCCGGAGCTCAGGCT 64  
 |||||  
 1 CCGACGTGGAAGAGAGAGAGAACTGCGCCCTGCTCATCCGGAGCTCAGGCT 60  
 |||||  
 65 CCAAGTGGGAGGCGACAAGATGTTCTCCCTCAAGAGTGAACGCGGTGCCATGTGA 124  
 |||||  
 61 CCAAGTGGGAGGCGACAAGATGTTCTCCCTCAAGAGTGAACGCGGTGCCATGTGA 120  
 |||||  
 125 GCTGGAGCTGGAGTGGATACGCGGCCAATCTCAGAGGTCCAGGTATGATGCTGTC 184  
 |||||  
 121 GCTGGAGCTGGAGTGGATACGCGGCCAATCTCAGAGGTCCAGGTATGATGCTGTC 180  
 |||||  
 185 TTAGATGTCAAGTGAAGAAACAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTATC 244  
 |||||  
 181 TTAGATGTCAAGTGAAGAAACAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTATC 240  
 |||||  
 245 ATTCTTCCACAATGCTGCTGATGCTGTGGTGGTGAACAGAAACAATGCTGCTCTCT 304  
 |||||  
 241 ATTCTTCCACAATGCTGCTGATGCTGTGGTGGTGAACAGAAACAATGCTGCTCTCT 300  
 |||||  
 305 GCCAGCGAGACTGGGTGTCCAAAGATCGGCAATGAGAGTGTGTAAGGCTTTTAG 364  
 |||||

Dp		301	GCCAGCAGACGACTGGGTGGTCCAAAAGAATCGGCAAAATGAGAGTGCTTAAGAAGCTTCCTTAG	360
Oy		365	CGCAATTGTTCAAGACCCTCGTGGATCTTTGTAATCACTAGTGCCTTACAAAAGCTGAACAAC	424
Dp		361	CGCAATTGTTCAAGACCCTCGTGGATCTTTGTAATCACTAGTGCCTTACAAAAGCTGAACAAC	420
Oy		425	TACAGGGGATGAATTCCTCAAAATGAGGCCCATGGAATCTGGTGC - TTTGGACTCATCA	482
Dp		421	TACAGGGGATGAATTCCTCAAAATGAGGCCCATGGAATCTGGTGCCTTTGGACTCATCA	480
Oy		483	AAGCCTTGTT - AGCATTTGTCAATTCTTATCTTACAGAAATTCCTGTGATTAAGAAGAT	540
Dp		481	AAGCCTTGTTAGATATTTGTGCATTTTATCTTCAGAAATTCCTGTGATTAAGAAGAT	540
Oy		541	AATTATTAAAGTGTGCTTCCTTCACTCTGTGTGTGTGTCGCGCACACAGCTTAGAAG	600
Dp		541	AATTATTAAAGTGTGCTTCCTTCACTCTGTGTGTGTGTCGCGCACACAGCTTAGAAG	600
Oy		601	TGCTATAAAAAAGAAAGAGCTCCCAATTGAATCAC - TTATTAATTACCATTTCTATA	659
Dp		601	TGCTATAAAAAAGAAAGAGCTCCCAATTGAATCACCTTTAATTATCCCATTTCTATA	660
Oy		660	CACACGACGATGGAAGCAGCTTTC - GAGACTTTTTCGATGCTTAG	703
Dp		661	CACACGACGATGGAAGCAGCTTTCGAGAACCTTTTTCATGCTTAG	706
<b>RESULT 14</b>				
B1226556		724 bp	mRNA	linear EST 11-JUL-2001
Locus		60295152761 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5095687 5', mRNA sequence.		
Accession		B1226556		
Version		B126556.1 GI:14680000		
Keywords		EST.		
Source		Homo sapiens (human)		
Organism		Homo sapiens		
Reference		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Authors		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
Title		1 (bases 1 to 724)		
Journal		NIH-MGC http://mgs.cni.nih.gov/.		
Comment		National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished		
		Contact: Robert Strausberg, Ph.D.		
		Email: csgrabs@e-mail.nih.gov		
		Tissue Procurement: Louis M. Staudt, M.D., Ph.D.		
		cDNA Library Preparation: Ling Hong/Rubin Laboratory		
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
		DNA Sequencing by: Incyte Genomics, Inc.		
		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:		
		http://image.lnl.gov		
		Plate: LNCMI866 row: f column: 08		
		High quality sequence stop:720.		

FEATURES	Location/Qualifiers	source
	1..724	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/dd_xref="taxon:9606"	
	/clone_image="5095687"	
	/clone_type="Burkitt lymphoma"	
	/lab_host="DH10 (phage-resistant)"	
	/clone_lib="NIH MGC 8"	
	/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5'-adaptor: GGCAcGAG (5'). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."	
BASE COUNT	195 a	
ORIGIN	153 c 189 g 187 t	

Query Match	84.1%;	Score 634.4;	DB 12;	Length 724;
Best Local Similarity	96.6%;	Pred. No. 8.4e-168;		
Matches 705; Conservative	0;	Mismatches 11;	Indels 14;	Gaps 5;

OY	18	CGGAGAGGAAACCTGCCCCCTGAGCTCTCACTCCGGAGAGCTCAGAGCTCAAGTCGGAGAG	77
Db	2	CGGAGAGGAAACCTGCCCCCTGAGCTCTCACTCCGGAGAGCTCAGAGCTCAAGTCGGAGAG	61
OY	78	CGACAAGATGTTCTCCCTCAAGAAAGTGGAACGCGGTGGCCATGTGAGAGCTGGGACGTGGA	137
Db	62	CGACAAGATGTTCTCCCTCAAGAAAGTGGAACGCGGTGGCCATGTGAGAGCTGGGACGTGGA	121
OY	138	GTGGGATATCGTGGCCCATCTTGCAGAGGATCCAGGTATGATATGCCCTGTGTAGATGTCAAG	197
Db	122	GTGGGATATCGTGGCCCATCTTGCAGAGGATCCAGGTATGATATGCCCTGTGTAGATGTCAAG	174
OY	198	TGAAGAACAAACAAAGAGACTGTGTGTGTGTGTGGGAGAAATGTATCATTTCTTCCACAA	257
Db	175	TGAAGAACAAACAAAGAGAGCTGTGTGTGTGTGTGGGAGAAATGTATCATTTCTTCCACAA	234
OY	258	CTGCTGCATGTCCCTGTGGGTGAAACAAACAATCGTGCCTCTCTGCCAGCAGACTG	317
Db	235	CTGCTGCATGTCCCTGTGGGTGAAACAAACAATCGTGCCTCTCTGCCAGCAGACTG	294
OY	318	GTGTGTCCAAAGATTCGGCAATATGAGTGTGTGAAGGCTTCTTAAGCCAGTTGTTCAAG	377
Db	295	GTGTGTCCAAAGATTCGGCAATATGAGTGTGTGAAGGCTTCTTAAGCCAGTTGTTCAAG	354
OY	378	AGCCCTGTGGATTTTGTAAATCCAGTCCCTACAAAGGCTTAGAACACTACAGGGAGTGA	437
Db	355	AGCCCTGTGGATTTTGTAAATCCAGTCCCTACAAAGGCTTAGAACACTACAGGGAGTGA	414
OY	438	TTCTTCAAAATAGAGCCGATGATCTGTGGTC--TTTGAAGCTCATCAAGCCTTGTT--	493
Db	415	TTCTTCAAAATAGAGCCGATGATCTGTGGTCCTTTGGAGCTCATCAAGCCTTGTTTA	474
OY	494	AGCATTTGTCAGTTTATCTTCAGAAATTTCTGTGATTTAAGAAATATTTATTTAAAG	553
Db	475	GCATTTTGTCAGTTTATCTTCAGAAATTTCTGTGATTTAAGAAATATTTATTTAAAG	534
OY	554	TGTCCTTCTTCTTACCTCTGTGGTGTGTGTGGGACACAGGCTTAGAAGTGCATTTAAAAAG	613
Db	535	TGTCCTTCTTCTTACCTCTGTGGTGTGTGTGGGACACAGGCTTAGAAGTGCATTTAAAAAG	594
OY	614	GAAAGACTCCAAATTAATCAAC--TTAATAATTTACCATTTCTTATCAACAGGACAGTGG	672
Db	595	GAAAGACTCCAAATTAATCAACCTTTAATAATTTACCATTTCTTATCAACAGGACAGTGG	654
OY	673	AAGCAGTTTC--GAGACTTTTTCATGTCTTATGTGTATCAGTTAAAAAGATGTTACA	730
Db	655	AAGCAGTTTCAGAGACATTTTGTGATGCTTATGTGTATCAGTTAAAAAGATGTTACA	714
OY	731	GTACCAATA 740	
Db	715	GTACCAATA 724	

RESULT	15
LOCUS	BUI92094
DEFINITION	BUI92094 922 bp mRNA linear EST 04-SEP-2002
ACCESSION	AGNCOURT_7968580 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011203
VERSION	5' mRNA sequence.
KEYWORDS	BUI92094 BUI92094.1 GI:22706078
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo. 1 (bases 1 to 922) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS	
TITLE	



```

QY 210 AGAGACTGTGTGTGGTGGGAGATATATCAATCTTCCACAACTGTGATGTC 269
    |||||
Db 181 AGAGACTGTGTGTGTGGGAGATATATCAATCTTCCACAACTGTGATGTC 240
QY 270 CCTGTGGTGAACAGAAACATGCTGCCCTCTCTGCACAGAGCTGGTGGTCCAAAG 329
    |||||
Db 241 CCTGTGGTGAACAGAAACATGCTGCCCTCTCTGCACAGAGCTGGTGGTCCAAAG 300
QY 330 AATCGGCAATGAGTGTGTAGAGAGCTTCTTAGGCGAGTGTTCAGAGCCCTGTGGA 389
    |||||
Db 301 AATCGGCAATGAGTGTGTAGAGAGCTTCTTAGGCGAGTGTTCAGAGCCCTGTGGA 360
QY 390 TCTTGAATTCAGTGGCCCTACAAAGGCTAGACACTAGAGGGAGTGAATCTTCAATAG 449
    |||||
Db 361 TCTTGAATTCAGTGGCCCTACAAAGGCTAGACACTAGAGGGAGTGAATCTTCAATAG 420
QY 450 GAGCCGATGATCTGTGTGTC--TTTGACTCATCAAGCCTTGG--TTAGCATTTGTCACT 506
    |||||
Db 421 GAGCCGATGATCTGTGTGTCCTTTGGGACTCATCAAGCCTTGGTTTGAATTTGTCACT 480
QY 507 TTTATCTTCAGAAATCTCTGTGATTAAGAAATATTTATTAAGTGTCTTCTTAC 566
    |||||
Db 481 TTTATCTTCAGAAATCTCTGTGATTAAGAAATATTTATTAAGTGTCTTCTTAC 540
QY 567 CTCTGTGTGTGTGTGGCGACACAGCTTGAAGTGTATTAAGAAAGAAAGAGCTCCAA 626
    |||||
Db 541 CTCTGTGTGTGTGTGGCGACACAGCTTGAAGTGTATTAAGAAAGAAAGAGCTCCAA 600
QY 627 ATTGAATCAAC--TTATATTTTACCAATTTCTATACACA--GGCAGTGAAGCAGTTTC-- 682
    |||||
Db 601 ATTGAATCAACCTTTATATATTTACCAATTTCTATACACAAGGGCAGTGAAGCAGTTTCAG 660
QY 683 GAGACTTTTTCATCTTATGTTGTTGATCACTTAAAAAGAAATTTACAGTAACA 737
    |||||
Db 661 AGAATCTTTTTCATCTTATGTTGTTGATCACTT--AAAAAGAAATTTACCGTAACAA 714

RESULT 17
LOCUS BG037017 893 bp mRNA linear EST 24-JAN-2001
DEFINITION BG037017.1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4374483 5',
            mRNA sequence.
ACCESSION BG037017
VERSION BG037017.1 GI:12432822
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
            1 (bases 1 to 893)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs@mail.nih.gov
            Tissue Procurement: Niklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
            Toshitaki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: LLM10038 row: d column: 04
            High quality sequence stop: 741.
            Location/Qualifiers
                1..893
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4374483"
                /tissue_type="hypothalamus"

```

```

/!ab host="DH10B"
/clone lib="NIH_MGC_96"
/!note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTN-3',
size-selected for average insert size 2.3 kb and
normalized to 50x. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC library."
BASE COUNT      241 a      202 c      253 g      197 t
ORIGIN
Query Match      82.7%; Score 623.6; DB 10; Length 893;
Best Local Similarity 97.1%; Pred. No. 16-164;
Matches 678; Conservative 0; Mismatches 14; Indels 6; Gaps 4;
QY 1 ATGGCCGACGTGGAGACGAGAGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 60
    |||||
Db 29 ATGGCCGACGTGGAGACGAGAGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 88
QY 61 GGCTCCAACTCGGAGGCGCAAGATGTTCTCTCTCAAGAGTGAACCGGATGCGCATG 120
    |||||
Db 89 GGCTCCAACTCGGAGGCGCAAGATGTTCTCTCTCAAGAGTGAACCGGATGCGCATG 148
QY 121 TGGAGCTGGAGAGTGGAGTGCATACGTGCGCATCTGAGAGGTCCAGTATGATGCC 180
    |||||
Db 149 TGGAGCTGGAGAGTGGAGTGCATACGTGCGCATCTGAGAGGTCCAGTATGATGCC 208
QY 181 TGTCTTGAATGTCAGAGCTGAAACAAACAAAGAGAGCTGTGTGGTGGGAGAGATGT 240
    |||||
Db 209 TGTCTTGAATGTCAGAGCTGAAACAAACAAAGAGAGCTGTGTGGTGGGAGAGATGT 268
QY 241 AATCAATCTCTTCCAACTGCTGATGTCCTCTGTGGTGAACAGAAACATGCTGCCCT 300
    |||||
Db 269 AATCAATCTCTTCCAACTGCTGATGTCCTCTGTGGTGAACAGAAACATGCTGCCCT 328
QY 301 CTCTGCGCAGAGAGCTGGGTGCTCCAAAGATTCGGCAATGAGTGTGTTAAGGCTTC 360
    |||||
Db 329 CTCTGCGCAGAGAGCTGGGTGCTCCAAAGATTCGGCAATGAGTGTGTTAAGGCTTC 388
QY 361 TTAGGCGAGTGTTCAGAGCCCTGTGGATCTGTATATCACTGCGCTTCAAGGCTAGA 420
    |||||
Db 389 TTAGGCGAGTGTTCAGAGCCCTGTGGATCTGTATATCACTGCGCTTCAAGGCTAGA 448
QY 421 ACACCTACAGGGGATGATTTCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 478
    |||||
Db 449 ACACCTACAGGGGATGATTTCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTC 508
QY 479 ATCAAAAGCTTGTGTT--AGCATTTGTCACTTTTATCTTCAAGAAATTTCTGTGATTA 536
    |||||
Db 509 ATCAAAAGCTTGTGTTAGCATTTTGTCACTTTTATCTTCAAGAAATTTCTGTGATTA 568
QY 537 AATATATTATTA--AGGATGCTCTTCTTCACTCTGTGTGTGTGCGGACACAGCTT 595
    |||||
Db 569 AATATATTATTA--AGGATGCTCTTCTTCACTCTGTGTGTGTGCGGACACAGCTT 628
QY 596 AGAAGTGTATTAAGAAAGAGAGCTCCAAATTTGATCACT--TTATATTTTACCATTT 654
    |||||
Db 629 AGAAGTGTATTAAGAAAGAGAGCTCCAAATTTGATCACTTTTATATTTACCATTT 688
QY 655 CTATACACAGGCACTGGAAGCAGTTTTCAGACTTTT 692
    |||||
Db 689 CTATACACAGGCACTGGAAGCAGTTTTCAGACTTTT 726

```

```

RESULT 18
LOCUS BX387707/c 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX387707 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS01037YPI3 3-PRIME, mRNA sequence.
ACCESSION BX387707

```



```

VERSION      BX387707.1  GI:30463313
KEYWORDS     EST
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS      Li W.B., Gruber C., Jesse J. and Polyes D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: sequef@genoscope.cns.fr Web: www.genoscope.cns.fr
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen. This sequence belongs to sequence cluster
              2013.f For more information about this cluster, see
              http://www.genoscope.cns.fr/
              cgi-bin/cluster.cgi?seq=CS0A1037CH07NP1&cluster=2013.f. Contact :
              Feng Liang Email: fliang@lifetech.com URL:
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID : CS0A1037CH07NP1.
              Location/Qualifiers
FEATURES
  source
    1..1201
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS0D1037YPL3"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-colo(dtr)
    primer: Five prime end enriched, double-strand cDNA was
    digested with Not I and cloned into the Not I and EcoR V
    sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT
  283 a 249 c 234 g 299 t 136 others
Query Match      82.1%; Score 619.4; DB 13; Length 1201;
Best Local Similarity 89.6%; Pred. No. 1.8e-163;
Matches 678; Conservative 6; Mismatches 66; Indels 7; Gaps 4;
QY 1 ATGCGCAGTGGAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 60
DB 757 ATGCGCAGTGGAGACGAGAGAGAAACCTGCGCCCTCTCACTCCGGAGCTCA 698
QY 61 GGCTCCAGTGGAGAGCGCAAGATGTTCTCCCTCAAGAACTGGAAGCGCGTGCCTATG 120
DB 697 GGCTCCAGTGGAGAGCGCAAGATGTTCTCCCTCAAGAACTGGAAGCGCGTGCCTATG 638
QY 121 TGGAGCTGGAGAGTGCATGCGTCCGCGCATCTGCGAGGTCGAGGTGATGATGCC 180
DB 637 TGGAGCTGGAGAGTGCATGCGTCCGCGCATCTGCGAGGTCGAGGTGATGATGCC 578
QY 181 TGTCTTGAATGTCAGAGTGAACCAAGAGAGCTGTGTTGTCGTGGGAGATGT 240
DB 577 TGTCTTGAATGTCAGAGTGAACCAAGAGAGCTGTGTTGTCGTGGGAGATGT 518
QY 241 AATCATTCCTCCCAACCTGTCATGTCCTCTGTCGGTGAACCAACATTCGCTGCCT 300
DB 517 AATCATTCCTCCCAACCTGTCATGTCCTCTGTCGGTGAACCAACATTCGCTGCCT 458
QY 301 CTCTCCAGCAGAGCTGGGTGTCGAAGAATCGGCAATGAGTGTGTTGAAGGCTTC 360
DB 457 NTCTGCCAGCAGAGCTGGGTGTCGAAGAATCGGCAATGAGTGTGTTGAAGGCTTC 398
QY 361 TTAGCGCAGTGTTCAGAGCCCTGTGTGATCTTGAATCAAGTCCCTCAACAAGCTTGA 420
DB 397 TTAGCGCAGTGTTCAGAGCCCTGTGTGATCTTGAATCAAGTCCCTCAACAAGCTTGA 338
QY 421 ACACACAGGGGAGTAAATTTCAATAGAGCGCATGAGATCTGTGTC--TTTGGATCTC 478
DB 337 AACTACTAGGGGAGTAAATTTCAATAGAGCGCATGAGATCTGTGTC--TTTGGNATTC 278

```

```

QY 479 ATCAAGCCTTGTT--AGCATTTGTCAGTTTATCTTCAGAAATTCCTGTGATTAGA 536
DB 277 ATNAAAGCCTTGTTTACATTTTNTGATTTTATCTTCAGAAATTCCTGTGATTAGA 218
QY 537 AGATATTTTAAAGGTCCTCTTACCTCTGTCGTGTGTCGGCAGACAGCTTA 596
DB 217 AGATATTTTAAAGGTCCTCTTACCTCTGTCGTGTGTCGGCAGACAGCTTA 158
QY 597 GAAGTGTATTAAGGAGAGAGCTCCAAATGAATCAC--TTATATTTTACCATTTC 655
DB 157 GAAGTGTATTAAGGAGAGAGCTCCAAATGAATCAC--TTATATTTTACCATTTC 98
QY 656 TATACAAAGCAGTGGAGAGATTTT--GAGACTTTTTCATGCTTATGTTATCAT 713
DB 97 TATACTAAGCAGTGGAGAGATTTT--GAGACTTTTTCATGCTTATGTTATCAT 38
QY 714 TAAAAAGATGTTTACATTAACAATTAAGTCAGTT 750
DB 37 TAAAAAGATGTTTACATTAACAATTAAGTCAGTT 1

```

```

RESULT 19
LOCUS      B1858307 797 bp mRNA linear EST 10-OCT-2001
DEFINITION 6033840801 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5393131 5',
            mRNA sequence.
ACCESSION  B1858307
VERSION    B1858307.1 GI:15999054
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: gga@lelmail.nih.gov
            Tissue Procurement: DCTD/DP
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1AM1201 row: 0 column: 20
            High quality sequence stop: 718.
            Location/Qualifiers
            1..797
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5393131"
            /tissue_type="mammary adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_87"
            /note="Organ: Breast; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.383 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC library."
BASE COUNT  229 a 166 c 204 g 198 t
ORIGIN

```

```

Query Match      81.7%; Score 616.2; DB 12; Length 797;
Best Local Similarity 95.0%; Pred. No. 1.2e-162;
Matches 725; Conservative 0; Mismatches 23; Indels 15; Gaps 8;
QY 4 GCCGACGTGAAGACGAGAGGAAACCTGCGCCCTGCTTCACTCCGAGCTCAGGC 63
DB 1 GCCGACGTGAAGACGAGAGGAAACCTGCGCCCTGCTTCACTCCGAGCTCAGGC 60

```



QY 64 TCCAGTGGGAGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGTGGCCATGTG 123  
 DB 61 TCCAGTGGGAGCGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGTGGCCATGTG 120  
 QY 124 AGCTGGGACGTGAGTGCATACGTGGCCATCTGCAGGGTCCAGGTATGATGCTGT 183  
 DB 121 AGCTGGGACGTGAGTGCATACGTGGCCATCTGCAGGGTCCAGGTATGATGCTGT 180  
 QY 184 CTAGATGTCACCTGAAAAAACAAGAGACCTGTGTGTGGTGGGAGAAATGAT 243  
 DB 181 CTAGATGTCACCTGAAAAAACAAGAGACCTGTGTGTGTGGTGGGAGAAATGAT 240  
 QY 244 CATTCCTTCCACAAGCTGTCATGTCCTGTGGGTGAACAACAATCGCTCCCTTC 303  
 DB 241 CATTCCTTCCACAAGCTGTCATGTCCTGTGGGTGAACAACAATCGCTCCCTTC 300  
 QY 304 TGGCAGACGAGCTGGGTGTGTCACAAGATCGCAATAGAGTGTAGAGGCTTCTTA 363  
 DB 301 TGGCAGACGAGCTGGGTGTGTCACAAGATCGCAATAGAGTGTAGAGGCTTCTTA 360  
 QY 364 GGGCAGTGTTCAGAGCCCTGGTGGATCTTGTATCAAGCCCTCAAAAGGCTAGAAC 423  
 DB 361 GGGCAGTGTTCAGAGCCCTGGTGGATCTTGTATCAAGCCCTCAAAAGGCTAGAAC 420  
 QY 424 CTACAGGAGATGATTTCTTCAATAGAGCCGATGATCTGTGTC--TTTGACTCATC 481  
 DB 421 CTACAGGAGATGATTTCTTCAATAGAGCCGATGATGATCTGTGTC--TTTGACTCATC 480  
 QY 482 AAAGCTTGGTT--AGCATTTGTCAGTTTATCTTCAAGAAATCTCTGTGATTAGAGA 539  
 DB 481 AAAGCTTGGTTGATTTGATTTGTCAGTTTATCTTCAAGAAATCTCTGTGATTAGAGA 540  
 QY 540 TAATTTATTT--AAAGTGTGTCCTTCAACCTGTGGTGTGTGCGGACACAGCTTAA 598  
 DB 541 TAATTTATTTAAAGTGTGTCCTTCAACCTGTGGTGTGTGCGGACACAGCTTAA 600  
 QY 599 AGTGTATATAAAGGAA--GAGCTCCAATTAATCAAC--TTATATTTTACCATTCT 656  
 DB 601 AGTGTATATAAAGGAAAGAGAGCTCCAATTAATCAACCTTATATTTTACCATTCT 660  
 QY 657 ATTCACAGGACGTGAGAGAGTTTC--GAGCTTTTTCAGTCTTATGTTGATCATGTT 714  
 DB 661 AT--ACAGGACGTGAGAGAGTTTCAGAGAACTTTTTCATGCTTATGTTGATCATGTT 717  
 QY 715 AAA--AAAGATGTTACAGTACCAATTAAGTGCAGTTTAA 754  
 DB 718 AACACAGATGTTTACAGTTACCAATTAAGTGCAGTTTAA 760  
 RESULT 20  
 B0648608 983 bp mRNA linear EST 15-JUL-2002  
 LOCUS B0648608  
 DEFINITION AGENCOURT\_8298360 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6269639  
 5', mRNA sequence.  
 ACCESSION B0648608  
 VERSION B0648608.1 GI:21772780  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NIH-MGC http://mgi.nci.nih.gov/  
 1 (bases 1 to 983)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov  
 Plate: LNCM2443 row: h column: 24  
 High quality sequence stop: 563.  
 Location/Qualifiers  
 1..983  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6269639"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."  
 BASE COUNT 227 a 247 c 282 g 227 t  
 ORIGIN  
 Query Match 81.3%; Score 612.8; DB 13; Length 983;  
 Best local similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 713; Conservative 0; Mismatches 32; Indels 11; Gaps 7;  
 QY 1 ATGGCGAGCTGGAGACGAGAGAGAAACCTGCGCTTCTCACTCCGGAGCTCA 60  
 DB 16 ATGGCGAGCTGGAGAGACGAGAGAGAAACCTGCGCTTCTCACTCCGGAGCTCA 75  
 QY 61 GGCTCCAACTCGGGAGGCGCAACAAGATGTTCTCCCTCAAGAAGTGAACCGGTGGCCATG 120  
 DB 76 GGCTCCAACTCGGGAGGCGCAACAAGATGTTCTCCCTCAAGAAGTGAACCGGTGGCCATG 135  
 QY 121 TGGAGCTGGGAGAGTGGAGTGCATACGTGCGCATCTGCAAGGATCCAGGTGATGATGCC 180  
 DB 136 TGGAGCTGGGAGAGTGGAGTGCATACGTGCGCATCTGCAAGGATCCAGGTGATGATGCC 195  
 QY 181 TGTCTTAGATGTCAGCTGAAAAACAACAAGAGACTGTGTGTGTGCTGGGAGAGATGT 240  
 DB 196 TGTCTTAGATGTCAGCTGAAAAACAACAAGAGACTGTGTGTGTGCTGGGAGAGATGT 255  
 QY 241 AATCATTTCTTCCACAACCTGTCATGTCCTGTGGTGAACAACAAGATGCTTCCTCT 300  
 DB 256 AATCATTTCTTCCACAACCTGTCATGTCCTGTGGTGAACAACAAGATGCTTCCTCT 315  
 QY 301 CTCTGCACAGAGACTGGGTGTGTCACAAGATCGCAATGAGAGTGTAGAGGCTTC 360  
 DB 316 CTCTGCACAGAGACTGGGTGTGTCACAAGATCGCAATGAGAGTGTAGAGGCTTC 375  
 QY 361 TTAGGCGAGTTGTCAGAGCCCTGTGTGATCTTGTATATCCAGTGCCTTCAAAAGCTTAA 420  
 DB 376 TTAGGCGAGTTGTCAGAGCCCTGTGTGATCTTGTATATCCAGTGCCTTCAAAAGCTTAA 435  
 QY 421 ACACTACAGGGATGAATTTCTTCAATAGAGACCGCATGATCTGTGTGTC--TTTGAAGTC 478  
 DB 436 ACACTACAGGGATGAATTTCTTCAATAGAGACCGCATGATCTGTGTGTC--TTTGAAGTC 495  
 QY 479 ATCAAGGCTTGGTT--AGCATTTGTCAGTTTATCTTCAAGAAATCTCTGTGATTAGAGA 536  
 DB 496 ATCAAGGCTTGGTTGATTTTGTGATTTTATCTTCAAGAAATCTCTGTGATTAGAGA 555  
 QY 537 AGATATATTTTAAAGTGTGCTTCTCACTCTGTGTGTGTGTGCGCACACAGCTTA 596  
 DB 556 AGATATATTTTAAAGTGTGCTTCTCACTCTGTGTGTGTGTGCGCACACAGCTTA 615  
 QY 597 GAAGGCTATATAAAGGAAAGAG--CTCCAATTAATCAAC--TTATATTTTACCATT 654  
 DB 616 GAAGGCTATATAAAGGAAAGAG--CTCCAATTAATCAAC--TTATATTTTACCATT 675  
 QY 655 CTATACACA--GGCAGTGAAGCAGTTTC--GAGACTTTTTCAGATGCTTAT--GGTGTAT 709

```

Db      676 CTATCAACAGGGGAGTGGAGCAGTCCGAGAACTGTTGATGCTTAAAGGCTAT 735
Qy      710 CAGTAAAAAAGATGTTACAGTAAACAATAAAGTG 745
Db      736 CAGCTGAAAAAACATTGCTGCGCCCAATAAAGG 771

RESULT 21
LOCUS   BG753323
DEFINITION 856 bp mRNA linear EST 15-MAY-2001
          6023174091 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4875489 5',
          mRNA sequence.
ACCESSION  BG753323
VERSION    BG753323.1 GI:14063976
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 856)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
          National Institutes of Health, Mammalian Gene Collection (MGC)
          JOURNAL Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: L16M1756 row: 0 column: 10
          High quality sequence stop: 767.
          Location/Qualifiers
            1..856
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:4875489"
              /tissue_type="normal pigmented retinal epithelium"
              /lab_host="DH10B (phage-resistant)"
              /clone_1ib="NIH_MGC_43"
              /note="Organ: eye; Vector: pOT87; Site 1: XhoI; Site 2:
              EcoRI; cDNA made by oligo-dt priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5'
              adaptor: GGCACGAG(G). Library constructed by Ling Hong
              in the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies).
              Note: this is a NIH_MGC Library."
BASE COUNT  231 a 214 c 211 g 200 t

Query Match      80.8%; Score 609; DB 10; Length 856;
Best Local Similarity 94.1%; Pred. No. 1.3e-160;
Matches 712; Conservative 0; Mismatches 30; Indels 15; Gaps 7;

```

```

Db      198 TGTCTTAGATGTCAGCTGAAACAAACAGAGACTGTGTGCTGGGAGAAATGT 257
Qy      241 AATCATTCCTTCCACAACTGCTGCATGTCCCTGTGGGTGAACAGAAATGCTGCCCT 300
Db      258 AATCATTCCTTCCACAACTGCTGCATGTCCCTGTGGGTGAACAGAAATGCTGCCCT 317
Qy      301 CTCTGCCAGCAGACGTGGGTGTCCCAAGAAATCGGCAAAAGAGTGTGAAGGCTTC 360
Db      318 CTCTGCCAGCAGACGTGGGTGTCCCAAGAAATCGGCAAAAGAGTGTGAAGGCTTC 377
Qy      361 TTACGCGAGTGTTCACAGGCCCTGTGGATCTTGAATCCAGTCCCTACAAAGGCTAGA 420
Db      378 TTACGCGAGTGTTCACAGGCCCTGTGGATCTTGAATCCAGTCCCTACAAAGGCTAGA 437
Qy      421 ACACACAGGGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 478
Db      438 ACACACAGGGGATGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGAATC 497
Qy      479 ATCAAGGCTTGGTT--AGCATTTGTCAGTTTATCTTCGAATTCCTGTGATTTAGA 536
Db      498 ATCAAGGCTTGGTTAGCATTTTGTGAGTTTATCTTCGAATTCCTGTGATTTAGA 557
Qy      537 AGATAATTATTTAA--AGGTGTCCTTCTCACTCTGTGTGTGTGTGCGACACAGCTT 595
Db      558 AGATAATTATTTAAAGGTGTCCTTCTCACTCTGTGTGTGTGTGCGACACAGCTT 617
Qy      596 AGAAGTG-CTATTAATAAAGAGAGAGAGCTCCAAATTTGAATCAC--TTATATTATCCATT 653
Db      618 AGAAGTGCTATTAATAAAGAGAGAGAGCTCCAAATTTGAATCACCTTTAATTTCCATT 677
Qy      654 TCTTTACACAGGAGGAGGAGAGAGAGCTTTTGGAGCTTTTGTGATGTC-----TTATGTTG 707
Db      678 TCTTTACACAGGAGGAGGAGAGAGAGCTTTTGTGATGTC-----TTATGTTG 737
Qy      708 ATCAGTTAA--AAAAGATGTTACAGTACCAATTTAA 742
Db      738 ATCAGTTAACCAAGATTTGTTACAGTACCAATTTAA 774

RESULT 22
LOCUS   B1601470
DEFINITION 767 bp mRNA linear EST 07-SEP-2001
          60324905391 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5300845 5',
          mRNA sequence.
ACCESSION  B1601470
VERSION    B1601470.1 GI:15494409
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 767)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
          National Institutes of Health, Mammalian Gene Collection (MGC)
          JOURNAL Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
          cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
          Toshiyuki and Piero Carninci (RIKEN)
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: L16M1761 row: j column: 14
          High quality sequence stop: 701.
          Location/Qualifiers
            1..767
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5300845"
FEATURES
  source

```



QY 427 CAGGGGATGATTTCTTCAATAGAGCGGATGATCTGTGTC--TTTGACTCATCAAA 484  
 Db 342 CAGGGGATGATTTCTTCAATAGAGCGGATGATCTGTGTC--TTTGACTCATCAAA 283  
 QY 485 GCCTTGCTT--AGCATTTGTCAGTTTATCTTCAAGAAATCTCTGTGATTAAGAAATA 542  
 Db 282 GCCTTGCTTACATTTCTTCAATAGAGCGGATGATCTGTGTC--TTTGACTCATCAAA 223  
 QY 543 TTTATTAAGTGTGCTTCTTCAATAGAGCGGATGATCTGTGTC--TTTGACTCATCAAA 602  
 Db 222 TTTATTAAGTGTGCTTCTTCAATAGAGCGGATGATCTGTGTC--TTTGACTCATCAAA 163  
 QY 603 CTATTAAGAAAGAAAGAGCTCCCAATTAATGATCAGC--TTATTAATTAATTAATTAATTA 661  
 Db 162 CTATTAAGAAAGAAAGAGCTCCCAATTAATGATCAGC--TTATTAATTAATTAATTAATTA 103  
 QY 662 ACAGGAGTGAAGAGCTTTC--GAGACTTTTTCAGTGTGATGATGATGATGATGATGAT 719  
 Db 102 ACAGGAGTGAAGAGCTTTC--GAGACTTTTTCAGTGTGATGATGATGATGATGATGAT 43  
 QY 720 AGAATGTACAGTACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 754  
 Db 42 AGAATGTACAGTACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 8

RESULT 24 847 bp mRNA linear EST 15-MAY-2001  
 BG773490  
 LOCUS 602720212P1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4837274 5',  
 DEFINITION mRNA sequence.

ACCESSION BG773490  
 VERSION BG773490.1 GI:14084143  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 847)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabbs@mail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshitaki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: L14M10769 row: 9 column: 03  
 High quality sequence stop: 775.  
 Location/Qualifiers

FEATURES  
 source  
 1. 847  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4837274"  
 /lab\_host="DH10B"  
 /clone\_1b="NIH\_MGC\_97"  
 /note="Organ testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtagag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to 10^5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT  
 ORIGIN

199 a 205 c 242 g 201 t

Query Match 80.5%; Score 606.6; DB 12; Length 847;  
 Best Local Similarity 94.2%; Pred. No. 6.3e-160;  
 Matches 685; Conservative 0; Mismatches 34; Indels 8; Gaps 5;

QY 1 ATGGCCGACGTGGAAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 60  
 Db 122 ATGGCCGACGTGGAAAGACGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 181  
 QY 61 GGCTTCAGTGGAGGAGGACCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCATTG 120  
 Db 182 GGCTTCAGTGGAGGAGGACCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCATTG 241  
 QY 121 TGGAGCTGGAGCGTGGAGTGGATGCTGCTGGCCATCTGGAGGGTCCAGTGAATGACC 180  
 Db 242 TGGAGCTGGAGCGTGGAGTGGATGCTGCTGGCCATCTGGAGGGTCCAGTGAATGACC 301  
 QY 181 TGTCTTAAGTCAAGCTGAAGAAACAAACAGAGAGCTGTGTGTGTGTGTGTGTGTGTGT 240  
 Db 302 TGTCTTAAGTCAAGCTGAAGAAACAAACAGAGAGCTGTGTGTGTGTGTGTGTGTGTGT 361  
 QY 241 AATCATTCCTTCCACACTGCTGATGTCCTCTGTGGGTGAACAGAAATCGTGCCT 300  
 Db 362 AATCATTCCTTCCACACTGCTGATGTCCTCTGTGGGTGAACAGAAATCGTGCCT 421  
 QY 301 CTCTGCCAGAGAGACTGGGTGTGCTCCAAAGAAATCGGCAATGAGATGATTGAAGCTTC 360  
 Db 422 CTCTGCCAGAGAGACTGGGTGTGCTCCAAAGAAATCGGCAATGAGATGATTGAAGCTTC 481  
 QY 361 TTAGCCGAGTGTTCAGAGCCCTGCTGTGATCTTGAATTCAGTGGCCCTACAAAGGCTAGA 420  
 Db 482 TTAGCCGAGTGTTCAGAGCCCTGCTGTGATCTTGAATTCAGTGGCCCTACAAAGGCTAGA 541  
 QY 421 ACACACAGGAGGAGTAATCTTCAATTAAGAGCCGATGATGATGATGATGATGATGATGAT 478  
 Db 542 ACACACAGGAGGAGTAATCTTCAATTAAGAGCCGATGATGATGATGATGATGATGATGAT 601  
 QY 479 ATCAAGACCTTGGTT--AGCATTTGCACTTTTATCTTGAAGAAATCTCTGTGATTAAGA 536  
 Db 602 ATCAAGACCTTGGTTGCACTTTTATCTTGAAGAAATCTCTGTGATTAAGA 661  
 QY 537 AGATTAATTTATTAAGTGTGCTCTTCTCACTCTGTGTGTGTGTGTC--CGCACACAGCTT 595  
 Db 662 AGATTAATTTATTAAGTGTGCTCTTCTCACTCTGTGTGTGTGTGTC--CGCACACAGCTT 721  
 QY 596 AGAAGTGTATTAAGAAAGAAAGAGCTCCAAATTAATCACTTAATTAATTAATTAATTAATTA 655  
 Db 722 AGAAGTGTATTAAGAAAGAAAGAGCTCCAAATTAATCACTTAATTAATTAATTAATTAATTA 781  
 QY 656 TATTAACAAGCAGTGGAGAGGATTTGAGAGCTTTTCAATGCTTAATGATGATGATGATGAT 715  
 Db 782 TATTAACA--AAGGCAAGGAGAGGATTTGAGAGCTTTTCAATGCTTAATGATGATGATGAT 838  
 QY 716 AAAAAA 722  
 Db 839 AAAAAA 845

RESULT 25 748 bp mRNA linear EST 10-APR-2001  
 BG576768  
 LOCUS 602599091P1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:4707838 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG576768  
 VERSION BG576768.1 GI:13584421  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 748)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabbs@mail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshitaki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: L14M10769 row: 9 column: 03  
 High quality sequence stop: 775.  
 Location/Qualifiers

## JOURNAL

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: DCTD/BTP

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL0578 row: e column: 23

High quality sequence stop: 748.

## FEATURES

source

Location/Qualifiers

1..748

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4707838"

/tissue\_type="mammary adenocarcinoma, cell line"

/lab\_host="NIH MGC 87"

/note="Organ: breast; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC library."

BASE COUNT 201 a 156 c 187 g 203 t 1 others

## ORIGIN

Query Match

Best Local Similarity 80.4%; Score 606; DB 10; Length 748;

Matches 686; Conservative 0; Mismatches 16; Indels 11; Gaps 6;

50 CCGGAGCTCAGGCTCCAGTCCGGAGGCGCAAGATGTTCTCCCTCAAGAATGGAAG 109

5 CCGGAGCTCAGGCTCCAGTCCGGAGGCGCAAGATGTTCTCCCTCAAGAATGGAAG 64

110 CCGTGGCCATGTGAGCTGGAGCGTGAAGTGCATGCGGCATCTGAGGAGGTCAGG 169

65 CCGTGGCCATGTGAGCTGGAGCGTGAAGTGCATGCGGCATCTGAGGAGGTCAGG 124

170 TGATGATGCTCTCTTAGATGTCAGAGCTGAAAAACAAGAGAGCTGTGTGTGTCT 229

125 TGATGATGCTCTCTTAGATGTCAGAGCTGAAAAACAAGAGAGCTGTGTGTGTCT 184

230 GGGGAAATGTATCATCTCTTCCACAACATGCTGCATGTCCTCTGCGTGAACAGAA 289

185 GGGGAAATGTATCATCTCTTCCACAACATGCTGCATGTCCTCTGCGTGAACAGAA 244

290 ATGCTGCTCTCTCTGCGAGAGCTGGGTGCTCCAAAGATCGCAATGAGAGTGTGT 349

245 ATGCTGCTCTCTCTGCGAGAGCTGGGTGCTCCAAAGATCGCAATGAGAGTGTGT 304

350 TAGAAGCTCTTAGAGGCAAGTGTTCAGAGCCTGTGTGATCTTGTATCCAGTGCCT 409

305 TAGAAGCTCTTAGAGGCAAGTGTTCAGAGCCTGTGTGATCTTGTATCCAGTGCCT 364

410 CAAAGCTTAGAACAACAGAGGATGATTTCTCAATAGAGAGCCAGTGTGTGTGT 469

365 CAAAGCTTAGAACAACAGAGGATGATTTCTCAATAGAGAGCCAGTGTGTGTGT 424

470 --TTTGAAGCTCAAGAGCTGTGT--AGCATTTGTGAGTTTATCTTCAAGAAATTC 525

425 CTTTGGAGCTCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 484

526 TGTGATTAAGAAATATTTATTAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 585

485 TGTGATTAAGAAATATTTATTAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 544

586 CACAGAGCTTAAGAGT 643

545 CACAGAGCTTAAGAGT 604

Qy

644 TTACCATTTCTATACAAAGAGAGTGAAGAGTTC--GAGCTTTTGCATGCTTA 701

605 TTACCATTTCTAT--ACAGGAGTGAAGAGTTCAGAGAGTTCATTTTGCATGCTTA 661

Db

702 TGTGATGATGATTAAGAAATGTTACAGTAACAATTAAGTGCATTTAA 754

662 TGTGATGATGATTAAGAAATGTTACAGTAACAATTAAGTGCATTTAA 714

RESULT 26

CB990062

LOCUS

AGENCYCOURT 13902487 NIH MGC 147 Homo sapiens cDNA clone

IMAGE:30346117 5', mRNA sequence.

ACCESSION

CB990062

VERSION

CB990062.1

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 782)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

JOURNAL

AUTHORS

CONTACT: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Dr. Stefan Hansson

CDNA Library Preparation: Michael J. Brownstein (NHGRI) with help

and advice from Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: NDAM385 row: b column: 14

High quality sequence stop: 635.

## FEATURES

source

1..782

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:30346117"

/tissue\_type="Human Placenta"

/lab\_host="NIH MGC 147"

/note="Organ: Placenta; Vector: pBluescriptR; Site:1: all-XhoI; Site:2: BamHI; Oligo-dt primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5; This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: This is a NIH MGC library."

BASE COUNT 195 a 186 c 215 g 186 t

## ORIGIN

Query Match

Best Local Similarity 80.0%; Score 603.4; DB 14; Length 782;

Matches 650; Conservative 0; Mismatches 16; Indels 7; Gaps 3;

1 ATGGCCGAGCTGGAAGAGCGAGAGAACTGCGCTGCTTCACCTCCGGAGCTTA 60

84 ATGGCCGAGCTGGAAGAGCGAGAGAACTGCGCTGCTTCACCTCCGGAGCTTA 143

61 GGCTCAAGTGGGAGGCGAGCAAGATGTTCTCCCTCAAGAGTGAAGCGGGTGGCCATG 120

144 GGCTCAAGTGGGAGGCGAGCAAGATGTTCTCCCTCAAGAGTGAAGCGGGTGGCCATG 203

121 TGAAGCTGGAGAGTGGAGTGCATAGTGGCCATCTGCAGGAGTCAAGTGAATGCC 180

204 TGAAGCTGGAGAGTGGAGTGCATAGTGGCCATCTGCAGGAGTCAAGTGAATGCC 263



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 692)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, [bento-soaresuiowa.edu](mailto:bento-soaresuiowa.edu)  
 Seq primer: M13 FORWARD  
 POLYA=yes

FEATURES  
 source  
 1..692  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-EDO-lys-n-12-0-UI"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP EDO"  
 /note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP EDO is a CDNA library containing the following tissue(s): Chondrosarcoma cell line C55. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAAGGCT.  
 TAG LIB=UI-H-EDO  
 TAG TISSUE=chondrosarcoma  
 TAG\_SEQ=CGTCAAGGCT"  
 169 c 135 g 206 t 1 others

BASE COUNT 181 a 169 c 135 g 206 t 1 others  
 ORIGIN

Query Match 79.0%; Score 595.8; DB 14; Length 692;  
 Best Local Similarity 97.0%; Pred. No. 6.5e-157;  
 Matches 650; Conservative 0; Mismatches 13; Indels 7; Gaps 4;

QY 92 CCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAGCGTGAAGTGGATACGTGGC 151  
 DB 692 CCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAGCGTGAAGTGGATACGTGGC 633  
 QY 152 CCATTCGCGGGTCCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 211  
 DB 632 CCATTCGCGGGTCCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 573  
 QY 212 AGGACTGT 271  
 DB 572 AGGACTGT 513  
 QY 272 TGTGGGTGAAGAGCAATCGCTGCTCTGTCCAGCAGAGACTGGGTGTGTCAAGAA 331  
 DB 512 TGTGGGTGAAGAGCAATCGCTGCTCTGTCCAGCAGAGACTGGGTGTGTCAAGAA 453  
 QY 332 TCGGCAAAAGAGGTGTTAGAGGCTTCTTACGAGAGTTGTTCAAGCCCTGGTGGATC 391  
 DB 452 TCGGCAAAAGAGGTGTTAGAGGCTTCTTACGAGAGTTGTTCAAGCCCTGGTGGATC 393  
 QY 392 TTGTATCCAGTCCCTACAAAGCTAGAACCTACAGGAGATGATCTTCAATATAGA 451

DB 392 TTGTATCCAGTCCCTACAAAGCTAGAACCTACAGGAGATGATCTTCAATATAGA 333  
 QY 452 GCCGATGATCTGTGGTCT--TTTGACTATCAAGCTTGGTT--AGCATTTGCAATT 507  
 DB 332 GCCGATGATCTGTGGTCTCTTTGGAGCTATCAAACTTGGTTGATGCAATT 273  
 QY 508 TTATCTTCGAAATTTCTGTGATTAAGAAATTAATTTAAAGTGTCTTCTTACC 567  
 DB 272 TTATCTTCGAAATTTCTGTGATTAAGAAATTAATTTAAAGTGTCTTCTTACC 213  
 QY 568 TCTGT 627  
 DB 212 TCTGT 153  
 QY 628 TTGAATCACC-TTATATATTACCATTCTTATPACACAGGACGTGACCACTTTC- 684  
 DB 152 TTGAATCACCATTATATATTACCATTCTTATPACACAGGACGTGACCACTTTCAGAG 93  
 QY 685 GACTTTTGGAGCTTATGTTGATGATGATGATGATGATGATGATGATGATGATGATG 744  
 DB 92 AACTTTTGCATGCTTATGTTGATGATGATGATGATGATGATGATGATGATGATGATG 33  
 QY 745 GCAGTTTAA 754  
 DB 32 GCAGTTTAA 23

RESULT 29  
 BG761806 716 bp mRNA linear EST 15-MAY-2001  
 LOCUS 602718011f1 NIH\_MGC\_49 Homo sapiens CDNA clone IMAGE:4841475 5',  
 DEFINITION mRNA sequence.  
 BG761806  
 ACCESSION BG761806.1 GI:14072459  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 716)  
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DTF  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LNCM1674 row: f column: 04  
 High quality sequence strop: 713.

FEATURES  
 source  
 1..716  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4841475"  
 /tissue\_type="melanotic melanoma, high MDR (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 49"  
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:  
 GCCAGAG(C). Size-selected 350bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-CDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 library."



```

BASE COUNT      178 a      166 c      198 g      174 t
ORIGIN
Query Match      78.8%; Score 594.2; DB 12; Length 716;
Best Local Similarity 96.1%; Pred. No. 1.9e-156;
Matches 664; Conservative 0; Mismatches 18; Indels 9; Gaps 5;

QY 1 ATGGCCGAGCTGGAAAGACGAGAGGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 60
DB 18 ATGGCCGAGCTGGAAAGACGAGAGGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 77
QY 61 GGCCTCAAGCTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGGGCGCATG 120
DB 78 GGCCTCAAGCTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGGGCGCATG 137
QY 121 TGGAGCTGGGAGCGTGGAGTGCATGATGCGCATGCGAGGTCCAGGTGATGATGCC 180
DB 138 TGGAGCTGGGAGCGTGGAGTGCATGATGCGCATGCGAGGTCCAGGTGATGATGCC 197
QY 181 TGTCTTAGATGTCACACTGAAAAAACAAGAGATGTGTGTGTGTGTGGGAGATGT 240
DB 198 TGTCTTAGATGTCACACTGAAAAAACAAGAGATGTGTGTGTGTGTGGGAGATGT 257
QY 241 AATCATTTCTTCCACACACTGCTGCATGTCCTGCTGGGTGAAAACAGAACATGCTGCCCT 300
DB 258 AATCATTTCTTCCACACACTGCTGCATGTCCTGCTGGGTGAAAACAGAACATGCTGCCCT 317
QY 301 CTCTGCGACGACGACTGGGTGTGTCACAAAGATCGGCAATGAGAGTGTGTAAGAGCTTC 360
DB 318 CTCTGCGACGACGACTGGGTGTGTCACAAAGATCGGCAATGAGAGTGTGTAAGAGCTTC 377
QY 361 TTAGGCGAGTGTTCAGAGCCCTGCTGAGATCTTGAATCAGTGGCCCTCAAGGCTTAA 420
DB 378 TTAGGCGAGTGTTCAGAGCCCTGCTGAGATCTTGAATCAGTGGCCCTCAAGGCTTAA 437
QY 421 ACACTACAGAGGATGATTTCTTCAATAGAGACCGATGATCTGTGTC--TTTGGATC 478
DB 438 ACACTACAGAGGATGATTTCTTCAATAGAGACCGATGATCTGTGTC--TTTGGATC 497
QY 479 ATCAAAGCTTGTTGTT--AGCATTTGTCACTTTTATCTTCAAAATTTCTGTGATTAAGA 536
DB 498 ATCAAAGCTTGTTGTT--AGCATTTGTCACTTTTATCTTCAAAATTTCTGTGATTAAGA 557
QY 537 AAT--AATTTTAAAGGTGCTCTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 594
DB 558 AATCAATTTATTAACAGGTGCTCTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 617
QY 595 TAGAAGTGTATTAATA--AAGAAAGAGCTCAAAATTGATCACTT--ATAATTACCA 651
DB 618 TAGAAGTGTATTAATA--AAGAAAGAGCTCAAAATTGATCACTT--ATAATTACCA 677
QY 652 TTTCTATACACAGGACGTGGAAGCATTTTC 682
DB 678 TTTCTATACACAGGACGTGGAAGCATTTTC 708

RESULT 30
LOCUS      BG037022      840 bp      mRNA      linear      EST 24-JAN-2001.
DEFINITION 60287341F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4374443 5',
ACCESSION  BG037022
VERSION     BG037022.1 GI:12432833
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 840)
AUTHORS   NIH-MGC http://mgi.mgi.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.

```

```

FEATURES
source
    Email: cgaabs-remail.nih.gov
    Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
    cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
    Toshiyuki and Piero Carninci (RIKEN)
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLI)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/ILNLI at:
    http://image.llnl.gov
    Plate: ILNLI0038 row: b column: 12
    High quality sequence stop: 806.
    Location/Qualifiers
        1..840
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:4374443"
        /tissue_type="hypothalamus"
        /lab_host="DH10B"
        /clone_id="NIH_MGC_96"
        /note="Organ: brain; Vector: pBluescriptR (modified
        pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gctcag
        ); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3'
        size-selected for average insert size 2.3 kb and
        normalized to ROT 5. This is a primary library enriched
        for full-length clones and constructed using the
        Cap-trapper method (Carninci, in preparation). Library
        constructed by M. Brownstein (NIH/NHGRI, National
        Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      219 a      191 c      232 g      198 t
ORIGIN
Query Match      78.6%; Score 592.8; DB 10; Length 840;
Best Local Similarity 92.7%; Pred. No. 5e-156;
Matches 701; Conservative 0; Mismatches 42; Indels 13; Gaps 7;

QY 1 ATGGCCGACCGTGGAAAGACGAGAGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 60
DB 29 ATGGCCGACCGTGGAAAGACGAGAGAAACCTGCGCCCTGCTCACTCCGGAGCTCA 88
QY 61 GGCCTCAAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGCGTGGCCATG 120
DB 89 GGCCTCAAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGCGTGGCCATG 148
QY 121 TGGAGCTGGGAGCGTGGAGTGCATGCTGCGCATCTGCAAGGTCACAGTGAATGATGCC 180
DB 149 TGGAGCTGGGAGCGTGGAGTGCATGCTGCGCATCTGCAAGGTCACAGTGAATGATGCC 208
QY 181 TGTCTTAGATGTCACAGCTGAAAAACAAGAGACGTGTGTGTGTGTGGGAGAAATGT 240
DB 209 TGTCTTAGATGTCACAGCTGAAAAACAAGAGACGTGTGTGTGTGTGGGAGAAATGT 268
QY 241 AATCATTTCTTCCACACACTGCTGCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 269 AATCATTTCTTCCACACACTGCTGCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 328
QY 301 CTCTGCGACGACGACTGGGTGTGTCACAAAGATCGGCAATGAGAGTGTGTAAGAGCTTC 360
DB 329 CTCTGCGACGACGACTGGGTGTGTCACAAAGATCGGCAATGAGAGTGTGTAAGAGCTTC 388
QY 361 TTAGCGAGTGTTCACAGGCGCTGCTGATCTGTATCCAGTGCCTTACAAAGGCTTGA 420
DB 389 TTAGCGAGTGTTCACAGGCGCTGCTGATCTGTATCCAGTGCCTTACAAAGGCTTGA 448
QY 421 ACACTACAGGAGTGAATTTCTCAATAGAGCGATGATCTGTGTC--TTTGGATC 478
DB 449 ACACTACAGGAGTGAATTTCTCAATAGAGCGATGATCTGTGTC--TTTGGATC 508
QY 479 ATCAAAGCTTGG--TTAGCATTTGTCACTTTTATCTTCAAGAAATCTCTGTGATTAAGAA 537
DB 509 ATCAAAGCTTGG--TTAGCATTTGTCACTTTTATCTTCAAGAAATCTCTGTGATTAAGAA 568
QY 538 GATATATTA--TTAAAGTGTCTCTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595

```



Db 569 GATTAATTATTTAAAGTGTGCTCTCTCACTCTGTGGTGTGTGGCGACACACTTT 628

Qy 596 AGAAGTGCATATAAAAAAGGAAAGAGC--TCCAAATTGAATCACCATTATA----ATTACC 649

Db 629 AGAAGTGCATATAAAAAAGGAAAGAGCTCCAAATTGGAATCACCCTTTTAATTTTAAACC 688

Qy 650 CATTTCTATACAACAGGCGAGTGGAGCAAGTTTC--GAGACTTTTTCGATGCTATGTTG 707

Db 689 ATTTCCTATACAACAGGCGAGTGGAGCAAGTTTTCAGGAGAACTTTTGGCTGCTATGTTG 748

Qy 708 ATCAGTTAAAAAAGATGTTACAGTAPACAATAAAG 743

Db 749 ATCCAGTTAACAAAGAGTGTCCGTACACAATPAGG 784

RESULT	31
LOCUS	B1828930
DEFINITION	B1828930 827 bp mRNA linear EST 04-OCT-2001 60307509221 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5166790 5', mRNA sequence.
ACCESSION	B1828930
VERSION	B1828930.1 GI:15940480
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	(bases 1 to 827)
AUTHORS	NIH-MGC <a href="http://mgc.nhl.nih.gov/">http://mgc.nhl.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: L1AM11414 row: h column: 23  
High quality sequence start: 2  
High quality sequence stop: 679.

## FEATURES

**Source**

Source	Base Count
1..827	232 a
/organism="Homo sapiens"	183 c
/mol_type="rRNA"	227 g
/db_xref="taxon:9606"	184 t
/clone="IMAGE:5166790"	1 others
/tissue_type="medulla"	
/lab_host="DH10B"	
/clone.lib="NIH MGC 119"	
/note="Organ: brain; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb; insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (InnVitrogen). Research Genetics tracking code 013. Note this is a NIH MGC Library."	

Query Match	78.2%	Score	589.6	DB	12	Length	827
Best Local Similarity	97.5%	Pred. No.	44-155				
Matches	653	Conservative	0	Mismatches	9	Indels	8
				Gaps			5
Qy	1	ATGCGCCGACGTGGGAAGACGAGAGGAAACCTCGCCCTTGCTCTTACTCTCCGAGACTCA	60				
Db	13	ATGCGCCGACGTGGGAAGACGAGAGGAAACCTCGCCCTTGCTCTTCACTCCGAGACTCA	72				

QY	61	GGCTCCAAAGTCGGAGGGGACAAGATGTTCTCCCTCAAGAAAGTGGAAACCCGGTGGCCATG	120
Dp	73	GGCTCCAAAGTCGGAGGGGACAAGATGTTCTCCCTCAAGAAATGGAAACCGGGTGGCCATG	132
QY	121	TGAGACTGGGACGTGGAGTGGCGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGGCC	180
Dp	133	TGAGACTGGACGTGGAGTGGCGATACGTGGCCCATCTGAGGGTCCAGGTGATGATGCC	192
QY	181	TGCTTTAGATGTCAAAGTGAACAAACAAGAGACTGTGTTGTGTCTGGGGAGATGT	240
Dp	193	TGCTTTAGATGTCAAAGTGAACAAACAAGAGACTGTGTTGTGTCTGGGGAGATGT	252
QY	241	AATCAATTCCTTCCACCACTGGCTGCATGTCCCGTGGGTGAACAGAACAAATGGCTGCCCT	300
Dp	253	AATCAATTCCTTCCACCACTGGCTGCATGTCCCGTGGGTGAACAGAACAAATGGCTGCCCT	312
QY	301	CTCTGCCAGCAGACTGGGTGGTCCAAAGATCGGCAATGAGTGTGTTAAGAGCTTC	360
Dp	313	CTCTGCCAGCAGACTGGGTGGTCCAAAGATCGGCAATGAGTGTGTTAAGAGCTTC	372
QY	361	TTAAGCGAGTTGTTCAAGACCCCTGGTGGATCTTGTAAATCCAGTGGCCCTACAAAGCTAGA	420
Dp	373	TTAAGCGAGTTGTTCAAGACCCCTGGTGGATCTTGTAAATCCAGTGGCCCTACAAAGGCTAGA	432
QY	421	ACACTACAGGGAGTGAATTTCTTCAAAATAGAGCCAGATGGATCTGTGATC--TTTGGATC	478
Dp	433	ACACTACAGGGAGTGAATTTCTTCAAAATAGAGCCAGATGGATCTGTGATCCTTTGGGATC	492
QY	479	ATCAAAAGCTTGGTT--AGCATTTGTCACTTTTATCTTCAGAAATTTCTCTGTGATTAAGA	536
Dp	493	ATCAAAAGCTTGGTTTAGCATTTTATGATCTTTTATCTTCAGAAATTTCTCTGTGATTAAGA	552
QY	537	AGATAATTTTAAAGGTGATCTTCCACCGCTGTGGTGTGTGTGGGACACACAGCTTA	596
Dp	553	AGATAATTTTAAAGGTGATCTTCCACCGCTGTGGTGTGTGTGGGACACACAGCTTA	612
QY	597	GAAAGTCTATAAA--AAAGAAAGACTCC--AAATTGATCAC--TTATAATTTTAAACCAT	652
Dp	613	GAAAGTCTATAAAACACAGAAAGAGCTCCAAAATTGATATCCTTTATAATTTAACCAT	672
QY	653	TTCTATACAA 662	
Dp	673	TTCTATACAA 682	

RESULT	32
LOCUS	B1858784
DEFINITION	B1858784 675 bp. mRNA linear EST 10-OCT-2001 603386387F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395388 5', mRNA sequence.

ACCESSION	BI858784	GI:15999531
VERSION	BI858784.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (pages 1 to 675)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph. D.

Email: [CGABs-r@mail.nih.gov](mailto:CGABs-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.llnl.gov>  
Plate: LLNL2007 row: m column: 21  
High quality sequence stop: 675.

## FEATURES

Location/Qualifiers

source

1. 675

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5395388"

/tissue\_type="mammary adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH\_MGC\_87"

/note="Organ: breast; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC library."

BASE COUNT 174 a 151 c 184 g 166 t

ORIGIN

Query Match 78.1%; Score 588.8; DB 12; Length 675;

Best Local Similarity 97.9%; Pred. No. 6.1e-155;

Matches 652; Conservative 0; Mismatches 2; Indels 12; Gaps 5;

QY 1 ATGGCCGACGTGGAGACGAGAGAGAAACCTGCGCTGCTCCTCCTCAGCTCCGGAGCTCA 60

DB 8 ATGGCCGACGTGGAGAGAGAGAGAGAACTGCGCTGCTCCTCCTCAGCTCCGGAGCTCA 66

QY 61 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAGAGAGTGAAGCGCGGTGCCATG 120

DB 67 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAGAGAGTGAAGCGCGGTGCCATG 126

QY 121 TGGAGCTGGAGACGTGGAGTGCATGTCGCGGCATCTGCGAGGCTCAGGTGATGATGCC 180

DB 127 TGGAGCTGGAGACGTGGAGTGCATGTCGCGGCATCTGCGAGGCTCAGGTGATGATGCC 179

QY 181 TGTCTTAGATGTCAAGCTGAAAAACAAGAGACTGTGTGTGTGTGTGTGGAGAGATGT 240

DB 180 TGTCTTAGATGTCAAGCTGAAAAACAAGAGACTGTGTGTGTGTGTGTGGAGAGATGT 239

QY 241 AATCATTTCTTCCCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

DB 240 AATCATTTCTTCCCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299

QY 301 CTCTCCAGCAGAGACTGGGTGTGTCAGAAAGATCGGCAATGAGAGTGTAGAGAGCTTC 360

DB 300 CTCTCCAGCAGAGACTGGGTGTGTCAGAAAGATCGGCAATGAGAGTGTAGAGAGCTTC 359

QY 361 TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGCCTTACAAAGGCTTGA 420

DB 360 TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGCCTTACAAAGGCTTGA 419

QY 421 AACATACAGGGAGTGAATTTCTCAATAGAGAGCCGATGATCTGTGTGCTCCTTGGAGCTC 478

DB 420 AACATACAGGGAGTGAATTTCTCAATAGAGAGCCGATGATCTGTGTGCTCCTTGGAGCTC 479

QY 479 ATCAAGAGCTTGG- TTAGCATTTGTCAAGTTTATCTTCAAGAAATCTCTGATTAAGA 537

DB 480 ATCAAGAGCTTGGTTAGCATTTGTCAAGTTTATCTTCAAGAAATCTCTGATTAAGA 539

QY 538 GATAATTTATTAAGGTGTCTCTCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 597

DB 540 GATAATTTATTAAGGTGTCTCTCTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 599

QY 598 AAGTCTTAAAG 656

DB 600 AAGTCTTAAAG 659

QY 657 ATACAA 662

DB 660 ATACAA 665

RESULT 33

LOCUS BG111792 947 bp mRNA linear EST 30-JAN-2001

DEFINITION 602285379p1 NIH\_MGC\_86 Homo sapiens CDNA clone IMAGE:4372819 5',

ACCESSION mRNA sequence.

VERSION BG111792

KEYWORDS BG111792.1 GI:12605298

REMARKS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 947)

AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LLM1003 row: n column: 20

High quality sequence stop: 682.

FEATURES

source

1. 947

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4372819"

/tissue\_type="osteosarcoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH\_MGC\_86"

/note="Organ: bone; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC library."

BASE COUNT 228 a 241 c 267 g 210 t 1 others

ORIGIN

Query Match 77.9%; Score 587.2; DB 10; Length 947;

Best Local Similarity 97.3%; Pred. No. 2e-154;

Matches 640; Conservative 0; Mismatches 13; Indels 5; Gaps 4;

QY 1 ATGGCCGACGTGGAGACGAGAGAGAAACCTGCGCTGCTCCTCCTCAGCTCCGGAGCTCA 60

DB 54 ATGGCCGACGTGGAGAGAGAGAGAGAACTGCGCTGCTCCTCCTCAGCTCCGGAGCTCA 113

QY 61 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAGAGAGTGAAGCGCGGTGCCATG 120

DB 114 GGCTCCAAAGTCGGAGGCGCAAGATGTTCTCCCTCAGAGAGTGAAGCGCGGTGCCATG 173

QY 121 TGGAGCTGGAGACGTGGAGTGCATGTCGCGGCATCTGCGAGGCTCAGGTGATGATGCC 180

DB 174 TGGAGCTGGAGACGTGGAGTGCATGTCGCGGCATCTGCGAGGCTCAGGTGATGATGCC 233

QY 181 TGTCTTAGATGTCAAGCTGAAAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

DB 234 TGTCTTAGATGTCAAGCTGAAAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 293

QY 241 AATCATTTCTTCCCAACACTGCTGATGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300

DB 294 AATCATTTCTTCCCAACACTGCTGATGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353

QY 301 CTCTCCAGCAGAGACTGGGTGTGTCAGAAAGATCGGCAATGAGAGTGTGTGAAGGCTTC 360

DB 354 CTCTCCAGCAGAGACTGGGTGTGTCAGAAAGATCGGCAATGAGAGTGTGTGAAGGCTTC 413

QY 361 TTAGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGCCTTACAAAGGCTTGA 420

DB 414 TTAGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATCCAGTGCCTTACAAAGGCTTGA 473

QY 421 AACATACAGGGAGTGAATTTCTCAATAGAGAGCCGATGATCTGTGTGCTCCTTGGAGCTC 478

```

|||||
474 ACCTCAGGGGATGATCTTCAATAGAGCCGATGATCTGTGCTTGGGACTC 533
479 ATCAAGACCTTG- TTAGATTTGTGATTTTATCTTCAAAATCTCTGTGTTAAGA 537
534 ATCAAGACCTTGTTAGATTTGTGATTTTATCTTCAAAATCTCTGTGTTAAGA 593
538 GATATTTTAAAGGTGCTCTTCTCACTCTGTGTGTGTG- TGCGCACACAG-CTT 595
594 GATATTTTAAAGGTGCTCTTCTCACTCTGTGTGTGTGTG- TGCGCACACAG-CTT 653
596 AGAAGTCTATATAAAGGAGAGAGTCCAAATGATCACTTATATTTATCCACT 653
654 AGAAGTCTATATAAAGGAGAGAGTCCAAATGATCACTTATATTTATCCACT 711

```

RESULT 34  
BQ632594  
LOCUS  
DEFINITION 629 bp mRNA linear EST 02-JUL-2002  
112504.Y1 HR85 islet Homo sapiens cDNA clone IMAGE:6031086.5,  
similar to TR:Q9Y5M7 Q9Y5M7 RING FINGER PROTEIN. [1] ; mRNA  
sequence.

ACCESSION BQ632594  
VERSION BQ632594  
KEYWORDS  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 629)  
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,  
Lemshik, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hiller, L., Marra, M., Pape, D., Wylie, T., Martin, J., Biststein, A.,  
Schmitt, A., Theising, B., Ritey, E., Komko, I., Bennett, J., Cardenas,  
M., Gibbons, W., McCann, R., Cole, R., Tsagaris, H., Williams, T.,  
Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium

TITLE  
JOURNAL  
COMMENT Unpublished  
Other ESTs: 112504.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -40R from Gibco  
High quality sequence stop: 495.  
Location/Qualifiers

## FEATURES

source

1. 629  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6031086"  
/issue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_id="HR85 islet"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:  
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.  
Size-selected on agarose gel. Average insert size  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permut Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2592."

BASE COUNT  
ORIGIN

157 a 141 c 173 g 158 t

Query Match 77.2%; Score 582.4; DB 13; Length 629;  
Best Local Similarity 98.4%; Pred No. 3.8e-153;  
Matches 610; Conservative 0; Mismatches 6; Indels 4; Gaps 2;

```

24 GGAACCTGCGCCCTGCTCACTCCGGAGCTCAGGCTCAAGTCGGAGCCACA 83
9 GGAACCTGCGCCCTGCTCACTCCGGAGCTCAGGCTCAAGTCGGAGCCACA 68
84 GATGTTCTCCCAAGAGTGAAGCGGTGGCCATGTGAGCTGGAGCTGAGTGC 143
69 GATGTTCTCCCAAGAGTGAAGCGGTGGCCATGTGAGCTGGAGCTGAGTGC 128
144 TACGCGCCATCTGAGGCTCAGGTGATGATGCTGCTTATGATGATGATG 203
129 TACGCGCCATCTGAGGCTCAGGTGATGATGATGATGATGATGATGATG 188
204 CAACAGAGAGCTGTGTTGTGCTGGGAGATGATGATGATGATGATGATG 263
189 CAACAGAGAGCTGTGTTGTGCTGGGAGATGATGATGATGATGATGATG 248
264 CATGTCCTGTGGGAGAGACAGACATGCTGCTCTCCAGACAGACTGGTGT 323
249 CATGTCCTGTGGGAGAGACAGACATGCTGCTCTCCAGACAGACTGGTGT 308
324 CCAAGAGATCGGCAATGAGTGTGATGAGAGCTTCTTATGAGGCTTCT 383
309 CCAAGAGATCGGCAATGAGTGTGATGAGAGCTTCTTATGAGGCTTCT 368
384 GGTGATCTTGTATTCAGTGTGCTTCAAGAGCTGATGATGATGATGATG 443
369 GGTGATCTTGTATTCAGTGTGCTTCAAGAGCTGATGATGATGATGATG 428
444 AATGAGAGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
429 AATGAGAGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
500 TGTGATCTTGTATTCAGTGTGCTTCAAGAGCTGATGATGATGATGATG 559
489 TGTGATCTTGTATTCAGTGTGCTTCAAGAGCTGATGATGATGATGATG 548
560 TTCTCACTCTGTGTTGTGCTGGCAGACAGCTTGAAGTGAATGAAGAGA 619
549 TTCTCACTCTGTGTTGTGCTGGCAGACAGCTTGAAGTGAATGAAGAGA 608
620 GCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 639
609 GCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 628

```

RESULT 35  
BG121625  
LOCUS  
DEFINITION 745 bp mRNA linear EST 30-JAN-2001  
60251581F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:449693.5,  
mRNA sequence.

ACCESSION BG121625  
VERSION BG121625.1  
KEYWORDS  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 745)  
NIH-MGC http://mgi.nci.nih.gov/.  
TITLE  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLAM10234 row: a column: 22  
High quality sequence stop: 653.

## FEATURES

Source

```

/oranism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4449693"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DHL10B (phage-resistant)"
/clone_1b="NIH MGc 90a"
/note="Organ: liver; Vector: pCMV-SPORE6; Site: 1; NotI; Site 2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length cDNAs and constructed by Life Technologies Note: This is a NIH MGc Library."

```

BASE COUNT	196 a	157 c	209 g	183 f
ORIGIN				

Query Match	77.2%	Score 582.4	DB 10	Length 745
Best Local Similarity	93.5%	Pred No. 4.1e-153		
Matches 675, Conservative	0	Mismatches 36	Indels 11	Gaps 6

[illegible]

QY	721	GA	722
Db	730	GA	731

### RESULT 36

LOCUS	734 bp	mRNA	linear	EST 21-JUL-2000
DEFINITION	BE397308	601286640F01 NIH_MGC_8	Homo sapiens	CDNA clone IMAGE:3619424 5',
KEYWORDS	mRNA sequence.			
ACCESSION	BE397308			
VERSION	BE397308.1	GI:9342673		
SOURCE	EST.			
	Homo sapiens (human)			

## REFERENCE

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMPTON

COMMENT

CDNA library preparation: Jang Hong/Rubin Laboratory  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LHCW289 row: c column: 09  
High quality sequence stop: 667.

## FEATURES

**Source**

```

/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3619424"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NH MGC_8"
/notes="Organ: lymph; Vector: pOT81; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G) . Size-selected >500bp for average insert size 1.8kb. Library constructed by Bing Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match	77.2%	Score 581.8	DB 10	Length 734
Best Local Similarity	93.4%	Pred. No. 66	153	
Best Match 666	Conservative	0	Mismatches 32	Indels 15
				Gaps 5

QY	37	CTGCCTCTCACTCCGGAGCTCAGGCTCCAAAGTCGGAGGCGCAACAAGATGTTCTCCCTC	96
Db	1	CTGCCCTCTCACTCCGGAGCTCAGGCTCCAAAGTCGGAGGCGCAACAAGATGTTCTCCCTC	60
QY	97	AAGAAGTGAACCGCGATGCGCATGTGAGCTGGAGCTGAGATCGCATACCTGCGCATC	156
Db	61	AAGAAGTGAACCGCGATGCGCATGTGAGCTGGAGCTGAGATCGCATACCTGCGCATC	120
QY	157	TGCAGGGTCCAGGATGATGATGCTCTGTTAGATGTCGAAGCTGAACCAACCAAGAGGAC	216
Db	121	TGCAGGGTCCA-----GATGCTCTGTTAGATGTCGAAGCTGAACCAACCAAGAGGAC	173
QY	217	TGCTGTTGCTGCTGGGGAGAAATGTAATCATTCCTTCACAACCTGCTGATTCCTCTGGG	276
Db	174	TGCTGTTGCTGCTGGGGAGAAATGTAATCATTCCTTCACAACCTGCTGATTCCTCTGGG	233
QY	277	GTGAACAGAACCAATCGCTCCCTCTTGCACAGGACTGGATGTGCACCAAGATCGGC	336

Db 234 GTGAAACAGAACATCGCTGCCCTCTCTGCCACAGAGACTGGGTGTCNAAGAAATGCG 293  
 QY 337 AAATGAGAGTGTTAGAAAGCTTTAGCCCACTTTGTCAGACCCCTGGTGATCTTGT 396  
 Db 294 AAATGAGAGTGTTAGAAAGCTTTAGCCCACTTTGTCAGACCCCTGGTGATCTTGT 353  
 QY 397 ATCCAGTGCCTTCAAAAGGCTAGAACACTACAGAGGATGAATCTTCAAAATAGAGCCGA 456  
 Db 354 ATCCAGTGCCTTCAAAAGGCTAGAACACTACAGAGGATGAATCTTCAAAATAGAGCCGA 413  
 QY 457 TGGATCTGTGTC -TTTGGACTCATCAAGCCTTGTT -AGCATTTGTCACTTTATC 512  
 Db 414 TGGATCTGTGTCCTTTGGAGCTCATCAAGCCTTGTTAGCATTTTGTCACTTTATC 473  
 QY 513 TTCAGAAATCTCTGTGATTAAGAGATATTTATTAAGAGTGTCTTCTTCACTTCTGT 572  
 Db 474 TTCAGAAATCTCTGTGATTAAGAGATATTTATTAAGAGTGTCTTCTTCACTTCTGT 533  
 QY 573 GGTGTGTGTCGCCACACAGCTTAGAAGTCTATTAAGAAAGAGAGCTCCAAATTTGAA 632  
 Db 534 GGTGTGTGTCGCCACACAGCTTAGAAGTCTATTAAGAAAGAGAGCTCCAAATTTGAA 593  
 QY 633 TCACCT--TATAATTTACCATTTCTATACACAGGACGTGAGAGAGTTTGAG-ACT 688  
 Db 594 TCACCTTTATTAATTTACCATTTCTATACACAGGACGTGAGAGAGAGTTTGAGACT 653  
 QY 689 TTTTCATGCTTATGTTGATGATGATTAAGAAAGAGTTTACATTAATTA 741  
 Db 654 TTTTCATGCTTATGTTGATGATGATTAAGAAAGAGTTTCCGCTACCAATTA 706

RESULT 37  
 B1757881 864 bp mRNA linear EST 25-SEP-2001  
 DEFINITION .603030486F1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5200676 5',  
 mRNA sequence.  
 B1757881  
 EST. B1757881.1 GI:15749459  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.liml.gov  
 Plate: LIML1502 row: 1 column: 21  
 High quality sequence stop: 667.  
 Location/Qualifiers  
 1. 864  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5200676"  
 /lab\_host="DH10B"  
 /clone\_11b="NIH\_MGC\_114"  
 /note="Organ: brain; Vector: PCMV-SPORT6; Site 1: NotI;  
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 6  
 male brains, age range 23-27 yo. Library is oligo-dT  
 primed and directionally cloned (EcoRV site is destroyed  
 upon cloning). Average insert size 1.5 kb, insert size  
 range 1-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:  
 this is a NIH MGC Library."  
 BASE COUNT 221 a 207 c 241 g 195 t  
 ORIGIN

Query Match 76.3%; Score 575.2; DB 12; Length 864;  
 Best Local Similarity 92.3%; Pred. No. 4,7e-151;  
 Matches 707; Conservative 0; Mismatches 43; Indels 16; Gaps 9;

QY 1 ATGGCCGACGTGGAAGACGAGAGAAACCTGCGCCCTTCTCACTCCGGAGCTCA 60  
 Db 76 ATGGCCGACGTGGAAGACGAGAGAAACCTGCGCCCTTCTCACTCCGGAGCTCA 135  
 QY 61 GGCTCCAAATCGGGAGGCGCAAGATGTTCTCCCTCAAAAGTGAACCGCGTGGCAAG 120  
 Db 136 GGCTCCAAATCGGGAGGCGCAAGATGTTCTCCCTCAAAAGTGAACCGCGTGGCAAG 195  
 QY 121 TGGAGCTGGGAGAGTGGAGTGCATACGTCGCGCATCTGCAAGGATCAGATGATGCC 180  
 Db 196 TGGAGCTGGGAGAGTGGAGTGCATACGTCGCGCATCTGCAAGGATCAGATGATGCC 255  
 QY 181 TGTCTTAGATGTCAGCTGAAGAAACAAAGAGACCTGTGTTGTGCTCGGAGATGT 240  
 Db 256 TGTCTTAGATGTCAGCTGAAGAAACAAAGAGACCTGTGTTGTGCTCGGAGATGT 315  
 QY 241 AATCATCTCTTCCACAACTGTCGATGTCCTGTCGTCGTCGTCGTCGTCGTCGTC 300  
 Db 316 AATCATCTCTTCCACAACTGTCGATGTCCTGTCGTCGTCGTCGTCGTCGTCGTC 375  
 QY 301 CTCTGCCACAGAGACTGGTGTGTCAGAAAGATCGGCAATGAGAGTGTAGAGGCTTC 360  
 Db 376 CTCTGCCACAGAGACTGGTGTGTCAGAAAGATCGGCAATGAGAGTGTAGAGGCTTC 435  
 QY 361 TTAGCGCAGTTGTTAGAGCCCTGCTGTGATCTTGTATTCAGTGCCTTCAAAAGCTTGA 420  
 Db 436 TTAGCGCAGTTGTTAGAGCCCTGCTGTGATCTTGTATTCAGTGCCTTCAAAAGCTTGA 495  
 QY 421 ACACTACAGAGGATGAATCTTCAATAGAGACCGATGATCTGTGTC -TTTGGACATC 478  
 Db 496 ACACTACAGAGGATGAATCTTCAATAGAGACCGATGATCTGTGTC -TTTGGACATC 555  
 QY 479 ATCAAGCCTTGTTAGCAATTTGTCAAGTTTATCTTCAGAAA-TTCTGTGTATTAAGA 537  
 Db 556 ATCAAGCCTTGTTAGCAATTTGTGTGATTTATCTTCAGAAAAGTTCTGTGATTAAGA 615  
 QY 538 GATTAATTTAT-AAAAGTGTCTTCTCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 595  
 Db 616 GATTAATTTATTAAGGTGTGACCTTCTCTGAGTGTGTGTGTGTGTGTGTGTGTGTGT 675  
 QY 596 AAGAGTGTAT-AAAAAGAGAAAGCTCCAAA--TTGAATCACCTTAT--AATTACG 649  
 Db 676 AAGAGTGTATCAACAAGAGAGAGCTCCAAAATTTGAATCAAGCTTATTAATTAAGGC 735  
 QY 650 CATTTCTATACAAAGGAGCTG-GGAAGCAGTTTC---GAGACTTTTGTGATGCTTATGG 704  
 Db 736 CATTTCTATACAAAGGAGCTG-GGAAGCAGTTTC---GAGACTTTTGTGATGCTTATGG 795  
 QY 705 TGTATCAGTTAAAGAAATGTTTACATTAACAATTAAGTGCAGTT 750  
 Db 796 GTGATCAGTTAAAGAAATGTTTAAAGTACATTAACGTGCAGTT 841

RESULT 38  
 Bg716023 793 bp mRNA linear EST 08-MAY-2001  
 LOCUS .602677415F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4800170 5',  
 mRNA sequence.  
 Bg716023  
 DEFINITION Bg716023.1 GI:13995210  
 ACCESSION Bg716023  
 VERSION Bg716023.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
1 (bases 1 to 793)  
Mammalia: Eutheria: Primates: Catarrhini; Hominiidae; Homo.  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM10690 row: m column: 03  
High quality sequence stop: 789.  
Location/Qualifiers

FEATURES  
source

1..793  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4800170"  
/issue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_96"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.3 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 192 a 198 c 211 g 192 t  
ORIGIN

Query Match 75.2%; Score 574.8; DB 10; Length 793;  
Best Local Similarity 95.0%; Pred. No. 5,9e-151;  
Matches 684; Conservative 0; Mismatches 17; Indels 19; Gaps 8;

1 ATGGCCGACGTGGAGACGAGAGAAACCTGCGCCCTCTACTCCGGAGCTCA 60  
80 ATGGCCGACGTGGAGACGAGAGAAACCTGCGCCCTCTACTCCGGAGCTCA 139  
61 GGCTCCAACTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGTGCATG 120  
140 GGCTCCAACTCGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACGGGTGCATG 199  
121 TGGAGCTGGAGAGTGGAGTACGTGCGGCATCTGAGGCTCAGTGATGATGCC 180  
200 TGGAGCTGGAGAGTGGAGTACGTGCGGCATCTGAGGCTCAGTGATGATGCC 252  
181 TGTCTTAGATGTCAAGCTGAAAAACAACAAGAGACTGTGTGTGTGGGAGATGT 240  
253 TGTCTTAGATGTCAAGCTGAAAAACAACAAGAGACTGTGTGTGTGGGAGATGT 312  
241 AATCATTTCTTCCAACTGCTGCTGATGTCCTGTGGGTGAACAAGAACATGCTGCTCC 300  
313 AATCATTTCTTCCAACTGCTGCTGATGTCCTGTGGGTGAACAAGAACATGCTGCTCC 372  
301 CTCTCCAGCAGAGACGCGGTGTGTCACAAAGATCGCAATGAGAGTGTGAAGGCTTC 360  
373 CTCTCCAGCAGAGACGCGGTGTGTCACAAAGATCGCAATGAGAGTGTGAAGGCTTC 432  
351 TTAGGCGAGTTGTTCAGAGCCTGTGTGATCTTGTATTCAGTGCCTTACAAGGCTAGA 420  
433 TTAGGCGAGTTGTTCAGAGCCTGTGTGATCTTGTATTCAGTGCCTTACAAGGCTAGA 492  
421 ACACACAGGGGATGAATTTCTTCAATATGAGAGCCGATGATGTGTGTC--TTTGACTC 478  
493 ACACACAGGGGATGAATTTCTTCAATATGAGAGCCGATGATGTGTGTC--TTTGACTC 552

479 ATCAAGCCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATCTCTGTGATTAGA 536  
553 ATCAAGCCTTGCTTATGACATTTTGTGATTTATCTTCAGAAATCTCTGTGATTAGA 612  
537 A-GATTAATTTATTA-AGTGGTCTTCTTACCTCTGTGTGTGTGTGTGCGGACAGCT 594  
613 ACGATTAATTTATTAACCGGTGTCTTCTTACCTCTGTGTGTGTGTGCGGACAGCT 672  
595 TAGAAGTCTTAAAAAAGAAA--GAGTCCAAATGAATCACC--TTATATTTACC 650  
673 TAGAAGTCTTAAAAAAGAAA--GAGTCCAAATGAATCACC--TTATATTTACC 732  
651 ATTCTATACCAACAGGACGTGAGACAGTTTC--GAGACTTTTTCGATGTTATGTTGA 708  
733 ATTCTATACCAACAGGACGTGAGACAGTTTC--GAGACTTTTTCGATGTTATGTTGA 792

RESULT 39  
LOCUS BG714665 651 bp mRNA linear EST 08-MAY-2001  
DEFINITION 602677016P1 NIH\_MGC\_96 Homo sapiens CDNA clone IMAGE:4799834 5',  
mRNA sequence.  
ACCESSION BG714665  
VERSION BG714665.1 GI:13993596  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NIH-MGC <http://mgc.nci.nih.gov/>.  
1 (bases 1 to 651)  
Mammalia: Eutheria: Primates: Catarrhini; Hominiidae; Homo.  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM10689 row: o column: 03  
High quality sequence stop: 643.  
Location/Qualifiers

FEATURES  
source

1..651  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4799834"  
/issue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_96"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.3 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 157 a 148 c 190 g 156 t  
ORIGIN

Query Match 75.8%; Score 571.8; DB 10; Length 651;  
Best Local Similarity 98.1%; Pred. No. 3.8e-150;  
Matches 611; Conservative 0; Mismatches 7; Indels 5; Gaps 3;

1 ATGGCCGACGTGGAGACGAGAGAAACCTGCGCCCTCTACTCCGGAGCTCA 60

```

Db      29  ATGGCCGACGTGAGAGACGAGAGAAACCTGCGCCCTGCTCTCACTCCGGAGACTCA 88
Qy      61  GGCTCCAGTGGAGAGCCGACAGATGTTCTCCCTCAAGAGGAGAACCGGCGGACAG 120
Db      89  GGCTCCAGTGGAGAGCCGACAGATGTTCTCCCTCAAGAGGAGAACCGGCGGACAG 148
Qy     121  TGGAGCTGGAGACGTGAGTGCATACGTGCGCCATCTGAGGAGTCCAGGATGATGAGC 180
Db     149  TGGAGCTGGAGACGTGAGTGCATACGTGCGCCATCTGAGGAGTCCAGGATGATGAGC 208
Qy     181  TGTCTTAGATGTAACTGAAAAACAAGAGAGACTGTGTGTGTCTGGGGAGATGT 240
Db     209  TGTCTTAGATGTAACTGAAAAACAAGAGAGACTGTGTGTGTCTGGGGAGATGT 268
Qy     241  AATCATCTCTTCCAGAACTGTCATGTCCTGTGTGTGTAACAGAAACAATGCGTCCCT 300
Db     269  AATCATCTCTTCCAGAACTGTCATGTCCTGTGTGTGTAACAGAAACAATGCGTCCCT 328
Qy     301  CTCTGCGACAGACAGCTGGTGTCTCAAGAGATGCGAAATGAGAGTGTAGAGGCTTC 360
Db     329  CTCTGCGACAGACAGCTGGTGTCTCAAGAGATGCGAAATGAGAGTGTAGAGGCTTC 388
Qy     361  TTAGCCCAATGTTTTCAGAGCCCTGTGTGATCTTGTATCCAGTCCCTACAAAGGCTAG 420
Db     389  TTAGCCCAATGTTTTCAGAGCCCTGTGTGATCTTGTATCCAGTCCCTACAAAGGCTAG 448
Qy     421  AACCTACAGGGAGTGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACATC 478
Db     449  AACCTACAGGGAGTGAATCTTCAATAGAGCCGATGATCTGTGTC--TTTGACATC 508
Qy     479  ATCAAAAGCCTTGTT--AGCATTTGTCAATTTATCTTGAAGAAATCTCTGTGATTAAGA 536
Db     509  ATCAAAAGCCTTGTT--AGCATTTGTCAATTTATCTTGAAGAAATCTCTGTGATTAAGA 568
Qy     537  AGATAATTTATTAAG--GTTGTCTTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGT 595
Db     569  AGATAATTTATTAAG--GTTGTCTTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGT 628
Qy     596  AGAAGGCTATAAAGAAAGAAAG 618
Db     629  AGAAGGCTATAAAGAAAGAAAG 651

RESULT 40
BO632301/c      641 bp      mRNA      linear      EST 02-JUL-2002
LOCUS      1123C04.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6031086 3'
DEFINITION similar to TR:Q9Y5M7 Q9Y5M7 RING FINGER PROTEIN. [1] ;, mRNA
sequence.
ACCESSION BO632301
VERSION BO632301.1 GI:21683819
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 641)
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,
Lemishka, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hiller, L., Maira, M., Pape, D., Wylie, T., Martin, J., Blinstein, A.,
Schmitt, A., Theising, B., Rletter, B., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McGann, R., Cole, R., Tsagaris, R., Williams, T.,
Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

```

```

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6031086"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_id="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permut Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."
BASE COUNT 176 a 159 c 125 g 181 t
ORIGIN
Query Match 75.1%; Score 566.4; DB 13; Length 641;
Best Local Similarity 97.2%; Pred. No. 1.3e-148;
Matches 620; Conservative 0; Mismatches 11; Indels 7; Gaps 4;
Qy     124  AGCTGGAGCGGAGATGCGATCGCGCCATCTGACAGGTCAGATGATGATGATGATGATGAT 183
Db     641  AGCTGGAGCGGAGATGCGATCGCGCCATCTGACAGGTCAGATGATGATGATGATGATGAT 582
Qy     184  CTTAGATGTCAAGCTGAAAAACAAGAGAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 243
Db     581  CTTAGATGTCAAGCTGAAAAACAAGAGAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 522
Qy     244  CATTCCTTCCACAATCTGTCATGTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 303
Db     521  CATTCCTTCCACAATCTGTCATGTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 462
Qy     304  TGCCAGAGAGATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 363
Db     461  TGCCAGAGAGATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 402
Qy     461  GCGCAGTTGTCAGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 423
Db     401  GCGCAGTTGTCAGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 342
Qy     424  CTACAGGGAGATGATTTCTTAATAGAGCCGATGATCTGTGTGT--TTTGACTCATC 481
Db     341  CTACAGGGAGATGATTTCTTAATAGAGCCGATGATCTGTGTGT--TTTGACTCATC 282
Qy     482  AAAAGCTTGTT--AGCATTTGTCTGTTTATCTTCAAGAAATCTCTGTGTGTGTGTGTGTGT 539
Db     281  AAAAGCTTGTTTACATTTGTCTGTTTATCTTCAAGAAATCTCTGTGTGTGTGTGTGTGTGT 222
Qy     540  TAAATTATTAAGGTGTCTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 599
Db     221  TAAATTATTAAGGTGTCTCTTCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 162
Qy     600  GTGTATATAAAGAGAGAGCTCCAAATGATCAGC--TTATAATTTACCATTTCTAT 658
Db     161  GTGTATATAAAGAGAGAGCTCCAAATGATCAGC--TTATAATTTACCATTTCTAT 102
Qy     659  ACAAGAGCAGTGAAGAGAGTTTGAG--ACTTTTCAATGCTTATGTTGATGATGATGATGAT 716
Db     101  ACAAGAGCAGTGAAGAGAGTTTGAG--ACTTTTCAATGCTTATGTTGATGATGATGATGAT 42
Qy     717  AAAAGATGTTACAGTAAACAATAAGAGCATTAA 754

```

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 449.  
Location/Qualifiers



Db 41 AAAAGATGTACAGTAACAATAAAGTCAGTTTAA 4

RESULT 41 678 bp mRNA linear EST 20-SEP-2002  
 BUE01136  
 LOCUS AGENCOURT 10018944 NIH\_MGC 142 Homo sapiens cDNA clone  
 DEFINITION IMAGE:6495009 5', mRNA sequence.  
 ACCESSION BUE01136  
 BUE01136.1 GI:23252895  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 678)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: NCI  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LDCM2672 row: c column: 10  
 High quality sequence stop: 527.  
 Location/Qualifiers  
 1..678  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6495009"  
 /tissue\_type="mixed (pool of 40 RNAs)"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_id="NIH\_MGC\_142"  
 /note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcgcc);  
 Site 2: SfiI (ggccgcgcgcgc); Double-stranded cDNA was  
 prepared from a pool of 40 cell line polyA+ RNAs (bladder  
 - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -  
 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,  
 kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,  
 ovary - 4%, pharynx - 2.5%, prostate 4.3%, salivary  
 gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were  
 used in cloning as follows:  
 5'-AAGCAGTGTATCAACGACAGTGCATTCAGCCGG-3' and  
 5'-ATTCTAGAGCGGAGCGCGGCAGCATG-dT(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the >0.5 kb  
 size fraction (other fractions present in NIH\_MGC\_141).  
 Library created in the laboratory of M. Brownstein (NIH,  
 NIH). Note: this is a NIH\_MGC library."

BASE COUNT 185 a 129 c 160 g 178 t 26 others

ORIGIN

Query Match 74.2%; Score 559.4; DB 13; Length 678;  
 Best local Similarity 96.8%; Pred. No. 1.2e-146;  
 Matches 613; Conservative 0; Mismatches 13; Indels 7; Gaps 4;

129 GGACGTGAGTGCATGATCGGCGCATCTGCAGGGCCAGGTGATGATGCTGCTTAG 188  
 1 GGAAGTGAAGTGCATGATCGGCGCATCTGCAGGGCCAGGTGATGATGCTGCTTAG 60

189 ATGTCAAGCTGAAAACAACAAGAGACTGTGTGTGCTGGGAGATGTAATCATTC 248  
 61 ATGTCAAGCTGAAAACAACAAGAGACTGTGTGTGCTGGGAGATGTAATCATTC 120

249 CTTCCACAACCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 308  
 121 CTTCCACAACCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 180

QY 309 GCAGACTGGGTGCTCCAAAGATCGCAAAATAGAGTGTAGAGGCTTTAGCCCA 368  
 Db 181 GCAGACTGGGTGCTCCAAAGATCGCAAAATAGAGTGTAGAGGCTTTAGCCCA 240  
 QY 369 GTTGTTCAGAGCCCTGGTGGATCTTGTATACAGGCTCCAAAGGCTAGAACCTCA 428  
 Db 241 GTTGTTCAGAGCCCTGGTGGATCTTGTATACAGGCTCCAAAGGCTAGAACCTCA 300  
 QY 429 GGGATGAATTCCTCAATAGAGCCGATGATCTGTGTC- TTGAGACTATCAAGC 486  
 Db 301 GGGATGAATTCCTCAATAGAGCCGATGATCTGTGTC- TTGAGACTATCAAGC 360  
 QY 487 CTGGT- AGCATTTGCACTTTATTTATTCAGAAATCTGTGATTAAGATTAAT 544  
 Db 361 CTGGTATGACATTTTTCAGATTTATCTTCAGAAATCTGTGATTAAGATTAAT 420  
 QY 545 TATTAAGTGTGCT 604  
 Db 421 TATTAAGTGTGCT 480  
 QY 605 ATAAAAAGAAAGAGCTCCAAATGATCAC- TTATTAATTTACCATTTCTATACAC 663  
 Db 481 ATAAAAAGAAAGAGCTCCAAATGATCACCTTTATTAATTTACCATTTCTATACAC 540  
 QY 664 AGCAGTGAAGCACTTC- GAGACTTTTCATGCTTATGTTGATCAGTTAAAAAG 721  
 Db 541 AGCAGTGAAGCACTTCGAGAACTTTTTCATGCTTATGTTGATCAGTTAAAAAG 600  
 QY 722 AATGTACAGTACAAATTAAGTCAGTTTAA 754  
 Db 601 AATGTACAGTACAAATTAAGTCAGTTTAA 633

RESULT 42 946 bp mRNA linear EST 11-JUN-2003  
 CD556579  
 LOCUS AGENCOURT 14477072 NIH\_MGC\_179 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30392712 5', mRNA sequence.  
 ACCESSION CD556579  
 CD556579.1 GI:31582647  
 VERSION EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 946)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDMA68 row: h column: 01  
 High quality sequence start: 22  
 High quality sequence stop: 606.  
 Location/Qualifiers  
 1..946  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30392712"  
 /tissue\_type="Pituitary"  
 /lab\_host="DH10B-Ton A ( T1 and T5 phage resistances) "

FEATURES  
 source



```

/clone lib="NIH_MGC_179"
/Note="Organ: brain; Vector: PCMV-SPORTS.1; Site: 1; EcoRV
(destroyed); Site 2: NotI; Library is oligo-dt primed and
directionally cloned (EcoRV site is destroyed upon cloning
). Average insert size 1.1 kb. Library was constructed by
(Invitrogen). Note: this is a NIH_MGC Library."
BASE COUNT      221 a      262 c      236 g      227 t
ORIGIN

```

```

Query Match      74.2% Score 559.2; DB 14; Length 946;
Best Local Similarity 95.1%; Pred. No. 1.6e-146;
Matches 656; Conservative 0; Mismatches 18; Indels 16; Gaps 7;

```

```

QY 1 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 60
Db 75 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 134
QY 61 GGCTCCAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 120
Db 135 GGCTCCAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGCGTGGCCATG 194
QY 121 TGGAGCTGGGACGTGAGTGGCGATAGTGGCCATCTGCAAGGATGATGCC 180
Db 195 TGGAGCTGGGACGTGAGTGGCGATAGTGGCCATCTGCAAGGATGATGCC 247
QY 181 TGTCTTAGATGTCAGCTGAAACAAACAAGAGAGAGCTGTGTGTGGTGGAGATGT 240
Db 248 TGTCTTAGATGTCAGCTGAAACAAACAAGAGAGAGCTGTGTGTGGTGGAGATGT 307
QY 241 AATCATTTCTTCCACAACTGCTGATGCTCTGTGGGTGAACAGAACTGCTGCTCC 300
Db 308 AATCATTTCTTCCACAACTGCTGATGCTCTGTGGGTGAACAGAACTGCTGCTCC 367
QY 301 CTCTGCAGAGAGAGCTGGGTGCTCAAGAAATGGGCAAAAGAGAGTGTGAAGGCTTC 360
Db 368 CTCTGCAGAGAGAGCTGGGTGCTCAAGAAATGGGCAAAAGAGAGTGTGAAGGCTTC 427
QY 361 TTAGCCGACGTGTTTCAAGAGCCCTGTGATCTGTATCCAGAGCCCTACAAAGGCTCA 420
Db 428 TTAGCCGACGTGTTTCAAGAGCCCTGTGATCTGTATCCAGAGCCCTACAAAGGCTCA 487
QY 421 ACACCTACAGGGAGATGATTTCTCAATAGAGAGCGATGATCTGTGTGCTTC 478
Db 488 ACACCTACAGGGAGATGATTTCTCAATAGAGAGCGATGATCTGTGTGCTTC 547
QY 479 AATCAAGCCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATTTCTGTATTAGA 536
Db 548 AATCAAGCCTTGCTT--AGCATTTGTCAGTTTATCTTCAGAAATTTCTGTATTAGA 607
QY 537 AGATAATTATTAAGGTGTCCTTCTTACCTCTGTGAGTGTGTGGCGACA--CAGCTT 595
Db 608 AGATAATTATTAAGGTGTCCTTCTTACCTCTGTGAGTGTGTGGCGACA--CAGCTT 667
QY 596 AGAAGTCTATAA--AAAAGAAAAGAGCTCCAAATTTGATCACC--TTATATTATCCA- 651
Db 668 AGAAGTCTATAA--AAAAGAAAAGAGCTCCAAATTTGATCACC--TTATATTATCCA- 727
QY 652 TTTCTATACAAACGCGAGTGAAGCAGTTT 681
Db 728 TTTCTATACCAACGCGAGTGAAGCAGTTT 757

```

```

RESULT 43
LOCUS      CB217926      622 bp      mRNA      linear      EST 06-FEB-2003
DEFINITION NISC nb05910.y1 COGENE 6E MAN Homo sapiens cDNA clone IMAGE:579514
5', mRNA sequence.
ACCESSION  CB217926
VERSION    CB217926
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      N1 (bases 1 to 622)
JOURNAL    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap
COMMENT    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb@mail.nih.gov
            CDNA Library Preparation:
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
            DNA Sequencing Center (NISC)
            Sequencing Center (NISC)
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plates: LLN12898 row: M column: 19
            Seq primer: M13RPI reverse primer (ABI).
            Location/Qualifiers

```

# FEATURES

```

source
1..622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:579514"
/issue_type="mandible, pooled"
/dev_stage="embryo, 6 weeks postconception"
/lab_host="DH10B"
/clone_lib="COGENE 6E MAN"
/Note="Vector: pAMP1; cDNA primed using oligo-dt primer,
directionally cloned into UDO sites of pAMP1. Size
selected for insert sizes ranging from 0.2-2.0 kb.
Normalized to Cots. Primary library, non-amplified.
Library constructed by M. Lovett. For more information on
this library, please contact R. Tidwell (Washington
University) or visit the COGENE website at
http://hg.wustl.edu/COGENE/."
BASE COUNT      144 a      147 c      182 g      149 t
ORIGIN

```

```

Query Match      74.1% Score 558.8; DB 14; Length 622;
Best Local Similarity 98.2%; Pred. No. 1.7e-146;
Matches 587; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

```

```

QY 1 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 60
Db 25 ATGGCCGACGTGGAAGACGAGAGAGAAACCTGGCCCTCTCACTCCGGAGCTCA 84
QY 61 GGCTCCAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGGTGGCCATG 120
Db 85 GGCTCCAGTGGGAGGCGCAAGATGTTCTCCCTCAAGAGTGAACCGGTGGCCATG 144
QY 121 TGGAGCTGGGACGTGAGTGGCGATAGTGGCCATCTGCAAGGATGATGCC 180
Db 145 TGGAGCTGGGACGTGAGTGGCGATAGTGGCCATCTGCAAGGATGATGCC 204
QY 181 TGTCTTAGATGTCAGCTGAAACAAACAAGAGAGCTGTGTGTGGTGGGAGAGATGT 240
Db 205 TGTCTTAGATGTCAGCTGAAACAAACAAGAGAGCTGTGTGTGGTGGGAGAGATGT 264
QY 241 AATCATTTCTTCCACAACTGCTGATGCTCTGTGGGTGAACAGAACTGCTGCTCC 300
Db 265 AATCATTTCTTCCACAACTGCTGATGCTCTGTGGGTGAACAGAACTGCTGCTCC 324
QY 301 CTCTGCAGAGAGAGCTGGGTGCTCCAAAGATGCGCAATGAGTGTGAGAGGCTTC 360
Db 325 CTCTGCAGAGAGAGCTGGGTGCTCCAAAGATGCGCAATGAGTGTGAGAGGCTTC 384
QY 361 TTAGCCGACGTGTTTCAAGAGCCCTGTGATCTGTGATCCAGTGGCCCTACAAAGGCTCA 420
Db 385 TTAGCCGACGTGTTTCAAGAGCCCTGTGATCTGTGATCCAGTGGCCCTACAAAGGCTCA 444
QY 421 ACACCTACAGGGAGATGATTTCTCAATAGAGAGCGATGATCTGTGTGCTTC 478
Db 445 ACACCTACAGGGAGATGATTTCTCAATAGAGAGCGATGATCTGTGTGCTTC 504

```

Oy	479	ATCAAAAGCCTTGCTT- AGCATTTGTGAGTTTTTAACCTCAGAATAATCCCTGGATTAGA	536
Dd	505	ATCAAAAGCCTTGCTTAGCATTTGTGCAGTTTTATCTTCAGAAATCTCTCGATTAGA	564
Oy	537	AGATATATTTAATAAGGTGCTCCTTCTTACTCTGTGTGTGTGTGTGCGCACACAGCT	594
Dd	565	AGATATATTTAATAAGGTGCTCCTTCTTACTCTGTGTGTGTGTGTGCGCACACAGCT	622

  

Db	RESULT 44	EST 09-OCT-2002
BUT30720/c	BUT30720	656 bp mRNA linear EST
LOCUS	UI-E-CII-afo-o-15-0-UI.s1 UI-E-CII Homo sapiens cDNA clone	
DEFINITION	UI-E-CII-afo-o-15-0-UI 3', mRNA sequence.	
ACCESSION	BUT30720	
VERSION	BUT30720.1 GI:23654893	
KEYWORDS	EST.	
SOURCE	Homo sapiens	
ORGANISM	Homo sapiens (human)	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Euteria; Primates; Carnivora; Hominiidae; Homo.	
TITLE	1 (bases 1 to 656) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery	
JOURNAL	Genome Res. 6 (9), 791-806 (1996)	
MEDLINE	97044477	
PUBMED	8889548	
COMMENT	Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA Tel.: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu Tissue procurement: Dr. Gregg Hageman cDNA library preparation: Dr. M. Bento Soares, University of Iowa CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Gene Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: MJ3 FORWARD POLYA=Yes.	

  

FEATURES	Location/Qualifiers
Source	1..656
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="UI-E-CII-afo-o-15-0-UI"
	/tissue type="RPE and Choroid"
	/dev stage="adult"
	/lab_host="DH10B (life Technologies) (T1 phage resistant)"
	/clone_1lb="UI-E-CII"
	/note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CII is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCGA. This library was created for the program Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
	TAG_LIB=UI-E-CII
	TAG_TISSUE=RPE and Choroid
	TAG_SEQ=ACCTA"

BASE COUNT	175 a	157 c	125 g	198 t	1 others
ORIGIN					
Query Match	74.0%	Score 558.2	DB 13	Length 656	
Best Local Similarity	96.7%	Pred. No. 2.ee-146			
Matches 613	Conservative	0	Mismatches 14	Indels 7	Gaps 4
Qy	128	GGGACGTGAGTGGCATACGTGGCCATCTGCAGGGTCCAGGTGATGATGTCCTGCTTA	187		
Db	656	GGGACGTGAGTGGCATACGTGGCCATCTGCAGGTCCAGGTGATGATGTCCTGCTTA	597		
Qy	188	GATGTCAGCTGAAAACAAACAAGAGACGTGTGTGGTCTGGGAGATGTAATCAT	247		
Db	596	GATGTCAGCTGAAAACAAACAAGAGACGTGTGTGGTCTGGGAGATGTAATCAT	537		
Qy	248	CCTTCACAACTGCTGCATGTCCTGTGGGTGAAACGAAACAAATGCTGCTCTGCTC	307		
Db	536	CCTTCACAACTGCTGCATGTCCTGTGGGTGAAACGAAACAAATGCTGCTCTGCTC	477		
Qy	308	AGCAGGACGTGGTGTGTCGAAAGAAATGCGCAATGAGATGTTAGAAAGCTTTTACGCG	367		
Db	476	AGCAGGACGTGGTGTGTCGAAAGAAATGCGCAATGAGATGTTAGAAAGCTTTTACGCG	417		
Qy	368	AGTTGTCAGAGCCCTGGTGGATCTTGAATCCAGTCCCTACAAAGCTAGAACACTAC	427		
Db	416	AGTTGTCAGAGCCCTGGTGGATCTTGAATCCAGTCCCTACAAAGCTAGAACACTAC	357		
Qy	428	AGGGGATGAATCTTCAATATGAGAGCCGATGATCTGTGTC - TTGGACTCATCAAG	485		
Db	356	AGGGGATGAATCTTCAATATGAGAGCCGATGATCTGTGTC - TTGGACTCATCAAG	297		
Qy	486	CCTTGCTT - AGCAATTTGTCAGTTTATCTTCAGAAATCTGTGATTTAAGAAGATAT	543		
Db	296	CCTTGCTTATGACATTTTGTGCTTATCTTCAGAAATCTGTGATTTAAGAAGATAT	237		
Qy	544	TTATTAAAGTGTGCTTCTCACTCTGTGTGTGTGTGCGGACACAGCTTGAAGATGC	603		
Db	236	TTATTAAAGTGTGCTTCTCACTCTGTGTGTGTGTGCGGACACAGCTTGAAGATGC	177		
Qy	604	TTATAAAAAGAAAGAGTCCAAATTTGAATCAC - TTATAATTACCATTTTATACAA	662		
Db	176	TTATAAAAAGAAAGAGTCCAAATTTGAATCAC - TTATAATTACCATTTTATACAA	117		
Qy	663	CAGGACGTGAGACGATTTTC - GAGACTTTTTCAGTCTTATGTTGATGCTGTTAAAAA	720		
Db	116	CAGGACGTGAGACGATTTTCAGAGAACTTTTTCAGTCTTATGTTGATGCTGTTAAAAA	57		
Qy	721	GAATGTTACGTAAACAATAAAGTCAATTAA 754			
Db	56	GAATGTTACGTAAACAATAAAGTCAATTAA 23			
RESULT 45					
BF033587					
LOCUS	595 bp	mRNA	linear	EST 20-OCT-2000	
DEFINITION	60145354F1 NIH_MGC_66 Homo sapiens CDNA clone IMAGE:385511 5',				
ACCESSION	BF033587				
VERSION	BF033587.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 595)				
JOURNAL	NIH-MGC <a href="http://mgs.nci.nih.gov/">http://mgs.nci.nih.gov/</a> .				
COMMENT	Unpublished				
	Contact: Robert Strausberg, Ph.D.				
	Email: cga@bbs.f-mail.nih.gov				
	Tissue Procurement: DCTD/DP				
	CDNA Library Preparation: Life Technologies, Inc.				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM587 row: 0 column: 16  
 High quality sequence stop: 595.

## FEATURES

## Source

1. .595  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3857511"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 66"  
 /note="Organ: ovary; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 153 a 126 c 165 g 151 t  
 ORIGIN

Query Match 73.7%; Score 555.4; DB 10; Length 595;  
 Best Local Similarity 98.2%; Pred. No. 1.6e-145;  
 Matches 584; Conservative 0; Mismatches 6; Indels 5; Gaps 2;

46 CACTCCGGAGCTCAGGCTCCAGTGGGAGGCGACAGATCTTCCCTCCAGAAAGTGG 105  
 1 CACTCCGGAGCTCAGGCTCCAGTGGGAGGCGACAGATCTTCCCTCCAGAAAGTGG 60  
 106 AACGCGGTGGCCATGTGGAGCTGGAGCTGGAGTGCATGCGCCATCTCAGAGGTC 165  
 61 AACGCGGTGGCCATGTGGAGCTGGAGCTGGAGTGCATGCGCCATCTCAGAGGTC 120  
 166 CAGGTGATGATGCTCTGTCTTAAGTGCAGAGCTGAAACAAACAAGAGAGCTGTGGT 225  
 121 CAGGTGATGATGCTCTGTCTTAAGTGCAGAGCTGAAACAAACAAGAGAGCTGTGGT 180  
 226 GTCGCGGGAGAAATGATATCATTCCTCCACAACCTGCGCATGTCCTGTGGTGGAAACG 285  
 181 GTCGCGGGAGAAATGATATCATTCCTCCACAACCTGCGCATGTCCTGTGGTGGAAACG 240  
 286 AACATCGCTGCTCTCTCTCCAGCAGAGCTGGGTGGTCCAAAGAAATCGCAATGAGAG 345  
 241 AACATCGCTGCTCTCTCTCCAGCAGAGCTGGGTGGTCCAAAGAAATCGCAATGAGAG 300  
 346 TGGTTGAAGGCTTCTTAGCGCAGTTGTTCAAGCCCTGTGATCTTGAATCCAGTGC 405  
 301 TGGTTGAAGGCTTCTTAGCGCAGTTGTTCAAGCCCTGTGATCTTGAATCCAGTGC 360  
 406 CCTACAAAGGCTTGAACACTACAGGGAGTGAATTTCTCAATAAGAGCCGATGATCTGT 465  
 361 CCTACAAAGGCTTGAACACTACAGGGAGTGAATTTCTCAATAAGAGCCGATGATCTGT 420  
 466 GGTCTCTTGGACTCATCAAGCCTTGTT---AGCATTTGTCAAGTTTATCTTCAGAAA 520  
 421 GGTCTCTTGGACTCATCAAGCCTTGTT---AGCATTTGTCAAGTTTATCTTCAGAAA 480  
 521 TTTCTGTGTATTAAGAAAGATTAATTTAAGAGTGTGCTTCTTACCTCTGTGGTGGT 580  
 481 TTTCTGTGTATTAAGAAAGATTAATTTAAGAGTGTGCTTCTTACCTCTGTGGTGGT 540  
 581 TCGCGCACACAGCTTGAAGTGTATTAAGAAAGAGAAAGAGCTCCAAATTAATCA 635  
 541 TCGCGCACACAGCTTGAAGTGTATTAAGAAAGAGAAAGAGCTCCAAATTAATCA 595

RESULT 46  
 BG026092 781 bp mRNA linear EST 24-JAN-2001  
 LOCUS 602232228F1 NIH\_MGC\_86 Homo sapiens cDNA clone IMAGE:4387076 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG026092

VERSION BG026092.1 GI:12413350  
 EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE 1 (bases 1 to 781)  
 JOURNAL NIH-MGC <http://mgi.nci.nih.gov/>  
 COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10070 row: p column: 21  
 High quality sequence stop: 777.

## FEATURES

## Source

1. .781  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4387076"  
 /tissue\_type="osteosarcoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 86"  
 /note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.53 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH MGC library."  
 BASE COUNT 216 a 163 c 210 g 192 t  
 ORIGIN

Query Match 73.5%; Score 554.2; DB 10; Length 781;  
 Best Local Similarity 94.8%; Pred. No. 3.8e-145;  
 Matches 693; Conservative 0; Mismatches 23; Indels 15; Gaps 11;

24 GGAACCTCGGCGCTGCTCCCTCCTCAGCTCGGAGCTCAGGCTCCAACTCGGAGGCGACAA 83  
 26 GGAACCTCGGCGCTGCTCCCTCCTCAGCTCGGAGCTCAGGCTCCAACTCGGAGGCGACAA 85  
 84 GATGTTCTCCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAGCTGGAGTGCAG 143  
 86 GATGTTCTCCTCAAGAGTGAACGCGGTGGCCATGTGAGCTGGAGCTGGAGTGCAG 144  
 144 TACGTGCGGCATCTGAGGGTCCAGGTGATGATGCTCTTGAATGTCAGCTGAAGA 203  
 145 TACGTGCGGCATCTGAGGGTCCAGGTGATGATGCTCTTGAATGTCAGCTGAAGA 204  
 204 CAAACAAGAGAGCTGTGTTGGTGTGGGAGAAATGATCAATTCCTTCCACAACCTGCG 263  
 205 CAAACAAGAGAGCTGTGTTGGTGTGGGAGAAATGATCAATTCCTTCCACAACCTGCG 263  
 264 CATGTCCTGTGGGTGAACACAGAACTGCTCTCTCCAGCAGAGCTGGGTGGT 323  
 264 CATGTCCTGTGGGTGAACACAGAACTGCTCTCTCCAGCAGAGCTGGGTGGT 323  
 324 CCAAGAAATCGGCAATGAGAGTGTGAAGAGCTTCTTAAGGCGAGTTGTCAGAGCCCT 383  
 324 CCAAGAAATCGGCAATGAGAGTGTGAAGAGCTTCTTAAGGCGAGTTGTCAGAGCCCT 382  
 384 GGTGATCTTGTATTCAGTGCCTTACAAAGGCTGAACACTACAGGGAGTGAATTTCTTC 443  
 383 GGTGATCTTGTATTCAGTGCCTTACAAAGGCTGAACACTACAGGGAGTGAATTTCTTC 442  
 444 AATAGAGAGCCGATGATCTGTGTC---TTTGAATCATCAAGCTTGTGTT---AGCAT 499  
 443 AATAGAGAGCCGATGATCTGTGTC---TTTGAATCATCAAGCTTGTGTT---AGCAT 502





Query Match 73.3%; Score 552.8; DB 12; Length 654;  
 Best Local Similarity 97.0%; Pred. No. 8.8e-145;  
 Matches 607; Conservative 0; Mismatches 12; Indels 7; Gaps 4;

```

QY 136 GAGTCCGATACGTGGCCATCTGCAAGGTCAGGTCATGATGATGCTGTCTTAAAGTGTAA 195
DB 648 GAGGCGGATACGTGGCCATCTGCAAGGTCAGGTCATGATGATGCTGTCTTAAAGTGTAA 589
QY 196 GGTGAAAACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255
DB 588 GCTGAAAACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
QY 256 AACTGTGATGATGCTGTGGTGAAGCAAGCAATGCTGCTCTCTGCGAGAGAGAG 315
DB 528 AACTGTGATGATGCTGTGGTGAAGCAAGCAATGCTGCTCTCTGCGAGAGAGAG 469
QY 316 TGGGTGTCTCAAGAAATGCGCAATGAGAGTGTGTAAGAGGCTTCTTACCGAGTGTTC 375
DB 468 TGGGTGTCTCAAGAAATGCGCAATGAGAGTGTGTAAGAGGCTTCTTACCGAGTGTTC 409
QY 376 AGAGCCCTGATGATGCTGTGAATCCAGTCCCTACCAAGGCTAGAACTACAGGAGATG 435
DB 408 AGAGCCCTGATGATGCTGTGAATCCAGTCCCTACCAAGGCTAGAACTACAGGAGATG 349
QY 436 AATTTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
DB 348 AATTTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
QY 494 --AGCATTTGTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 551
DB 288 TAGCATTTTGTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 229
QY 552 GGTGTCTCTCTCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 611
DB 228 GGTGTCTCTCTCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 169
QY 612 AGGAAAGAGCTCCAAATTTGATGAC--TTATAATTTTACCATTCTTATCAACAGGAGT 670
DB 168 AGGAAAGAGCTCCAAATTTGATGAC--TTATAATTTTACCATTCTTATCAACAGGAGT 109
QY 671 GGAAGAGATTTT--GAGACTTTTGTGATGATGATGATGATGATGATGATGATGATGAT 728
DB 108 GGAAGAGATTTT--GAGACTTTTGTGATGATGATGATGATGATGATGATGATGATGAT 49
QY 729 CAGTAACAATAAAGTCAAGTTTAA 754
DB 48 CAGTAACAATAAAGTCAAGTTTAA 23

```

RESULT 50  
 BG718227 715 bp mRNA linear EST 08-MAY-2001  
 LOCUS 60266194F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4828198 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG718227.1 GI:13997414  
 VERSION BG718227.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 715)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Miklos Palokovics, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

Found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LLM10745 row: 1 column: 23  
 High quality sequence stop: 714.  
 Location/Qualifiers  
 1. 715

FEATURES  
 source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4828198"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescript (modified  
 pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTTT-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 151 a 184 c 217 g 163 t  
 ORIGIN

Query Match 72.4%; Score 546.2; DB 10; Length 715;  
 Best Local Similarity 98.0%; Pred. No. 6.7e-143;  
 Matches 575; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

```

QY 1 ATGSCCAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 129 ATGSCCAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
QY 61 GGTCTCAAGTGGGAGGCGACAGATGTTCTCTCAAGAGTGAACCGGGTGGCCATG 120
DB 189 GGTCTCAAGTGGGAGGCGACAGATGTTCTCTCAAGAGTGAACCGGGTGGCCATG 248
QY 121 TGGAGCTGGGAGCGTGAAGTGGATGAGCGCCATCTGCAAGGTCAGAGTATGATGCC 180
DB 249 TGGAGCTGGGAGCGTGAAGTGGATGAGCGCCATCTGCAAGGTCAGAGTATGATGCC 308
QY 181 TGTCTAGATGTCAAGCTGAAGCAACAAACAGAGAGAGTGTGTGTGTGTGTGTGTGTGT 240
DB 309 TGTCTAGATGTCAAGCTGAAGCAACAAACAGAGAGAGTGTGTGTGTGTGTGTGTGTGT 368
QY 241 AATCATTTCTTCCCAACTGCTGATGCTCTGTGGGTGAACAGAAATCGTGCCT 300
DB 369 AATCATTTCTTCCCAACTGCTGATGCTCTGTGGGTGAACAGAAATCGTGCCT 428
QY 301 CTGTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 429 CTGTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 361 TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATTCAGTGCCTCAAGAGCTAGA 420
DB 489 TTAGGCGAGTGTTCAGAGCCCTGTGTGATCTTGTATTCAGTGCCTCAAGAGCTAGA 548
QY 421 ACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 478
DB 549 ACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
QY 479 ATCAAGAGCTTGGTT--AGCATTTGTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 536
DB 609 ATCAAGAGCTTGGTT--AGCATTTGTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 668
QY 537 AGATATTTTATTAAGGTGTGTCTTCTCACTCTGTGTGTGTGTGTGTGTGTGTGTGT 583
DB 669 AGATATTTTATTAAGGTGTGTCTTCTCACTCTGTGTGTGTGTGTGTGTGTGTGTGT 715

```

Search completed: November 7, 2003, 15:15:50  
 Job time: 1912.92 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 11:11:37 ; Search time 823.044 Seconds

(without alignments)  
6163.455 Million cell updates/sec.

Title: US-09-509-779-3\_COPY\_141\_264

Sequence: 1 CGATACGTCGCGCATCTGCA.....ATTCTTCACACACGCTGTC 124

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pl:\*  
10: gb\_ro:\*  
11: gb\_ro:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pln:\*  
35: em\_hlg\_mam:\*  
36: em\_hlg\_rtd:\*  
37: em\_hlg\_vrt:\*  
38: em\_hgo\_hum:\*  
39: em\_hgo\_mus:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	100.0	342	6	AX282592 Sequence
2	124	100.0	342	9	AF142060 Homo sapi
3	124	100.0	342	9	BT007348 Homo sapi
4	124	100.0	754	6	BD096967 SAG:apopt
5	124	100.0	754	6	BD096989 SAG:apopt
6	124	100.0	754	6	BD096994 SAG:apopt
7	124	100.0	754	6	BD096995 SAG:apopt
8	124	100.0	754	6	AF092878 Homo sapi
9	124	100.0	816	9	BC005966 Homo sapi
10	124	100.0	822	9	BC005966 Homo sapi
11	124	100.0	836	9	AF164679 Homo sapi
12	121	97.6	754	6	BD096988 SAG:apopt
13	116	93.5	754	6	BD096981 SAG:apopt
14	112	90.3	754	6	BD096987 SAG:apopt
15	107	86.3	754	6	BD096982 SAG:apopt
16	107	86.3	754	6	BD096983 SAG:apopt
17	103	83.1	754	6	BD096986 SAG:apopt
18	103	83.1	754	6	BD096990 SAG:apopt
19	97	78.2	754	6	BD096985 SAG:apopt
20	90	72.6	747	6	BD096973 SAG:apopt
21	89	71.8	674	9	AF312226 Homo sapi
22	83	66.9	754	6	BD096984 SAG:apopt
23	76	61.3	754	6	BD096992 SAG:apopt
24	74	59.7	754	6	BD096991 SAG:apopt
25	74	59.7	754	6	BD096993 SAG:apopt
26	49	39.5	35638	9	AC112771 Homo sapi
27	49	39.5	125041	2	AC068693 Homo sapi
28	49	39.5	238330	2	AC106176 Rattus no
29	49	39.5	245476	2	AC098496 Rattus no
30	48	38.7	187353	2	AC118192 Mus muscu
31	43	34.7	324	6	BD077705 S'EST OF
32	43	34.7	59155	2	AC104983 Homo sapi
33	43	34.7	163521	2	AC108040 Homo sapi
34	43	34.7	164500	2	AC092418 Homo sapi
35	42	33.9	706	6	BD096974 SAG:apopt
36	36	29.8	254832	3	AC095698 Rattus no
37	36	29.0	120515	10	AF929565 Mouse DNA
38	35	28.2	439	9	AJ33208 Homo sapi
39	35	28.2	264606	2	AC137864 Rattus no
40	32	25.8	1125	10	BC011127 Mus muscu
41	32	25.8	1140	6	BD096986 SAG:apopt
42	32	25.8	1140	10	AF092877 Mus muscu
43	32	25.8	219370	10	AC123935 Mus muscu
44	31	25.0	612	11	BV071691 S21P6546
45	31	25.0	112321	2	AC103666 Mus muscu

#### ALIGNMENTS

RESULT 1  
AX282592 342 bp DNA linear PAT 02-NOV-2001  
LOCUS AX282592  
DEFINITION Sequence 7 from Patent WO0175145.  
ACCESSION AX282592  
VERSION AX282592.1 GI:16609675  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Issakani, S.D., Huang, J., Sheung, J. and Pray, T.R.  
TITLE Ubiquitin ligase assay  
JOURNAL Patent: WO 0175145-A 7 11-OCT-2001;

Pred. No. is the number of results predicted by chance to have a





	Matches	124;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CGATPACGTGCCCATCTGCGAGGTC	CAGGTCCAGGTGATGATGCTGTCCTTAAGATGTCAAGCTGA	60						
Db	141	CGATPACGTGCCCATCTGCGAGGTC	CAGGTCCAGGTGATGATGCTGTCCTTAAGATGTCAAGCTGA	200						
QY	61	AAACAAACAAGAGCATGTGTTGTGGTCTGGGGAGATGTAATCAATCTCCCAACTG	120							
Db	201	AAACAAACAAGAGCATGTGTTGTGGTCTGGGGAGATGTAATCAATCTCCCAACTG	260							
QY	121	CTGC	124							
Db	261	CTGC	264							

RESULT 4				
BD096967				
LOCUS	BD096967	754 bp	DNA	linear
DEFINITION	SAG:apoptosis sensitivity gene.			
ACCESSION	BD096967			
VERSION	BD096967.1	GI:22642555		
KEYWORDS	JP 2001526063-A/2.			
SOURCE	unidentified			
ORGANISM	unidentified			
REFERENCE	1 (bases 1 to 754)			
AUTHORS	Sun,Y.			
TITLE	SAG:apoptosis sensitivity gene			
JOURNAL	Patent: JP 2001526063-A 2 18-DEC-2001;			
	WARNER LAMBERT CO			
COMMENT	OS Unidentified			

PF	15-DEC-1998 JP	2000525451			
PR	19-DEC-1997 US	60/068119, 11-SEP-1998 US	60/099840	PI	
Y1	SUN				
PC	C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,				
PC	A61P39/06,				
PC	A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/21 PC				
	, C12N5/10, C12Q1/68,				
PC	G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC				
Strandedness:	Double;				
CC	Topology: Linear;				
CC	/note = 'Human SAG'				
FH	Key	Location/Qualifiers			
FT	CDS	1..339			
FT	mat_peptide	1..339			
FT	misc_feature	1..754.			
	Location/Qualifiers				
	1..754				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
BASE COUNT	205 a	155 c	201 g	193 t	
ORIGIN					

Query Match	Similarity	100.0%	Score 124	DB 6	Length 754
Best Local	Similarity	100.0%	Pred. No. 6.4e-66		
Matches	124	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
QY	1	CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTCTGTCTTAGATGTCAAGCTGA	60		
Db	141	CGATACGTGGCCATCTGCAGGGTCCAGGTGATGATGCTCTGTCTTAGATGTCAAGCTGA	200		
QY	61	AAACAAACAAAGAGACTGTGTGTGTCTGGGAGATGATCAATTCCTTCCAACTG	120		
Db	201	AAACAAACAAAGAGACTGTGTGTGTCTGGGAGATGATCAATTCCTTCCAACTG	260		
QY	121	CTGC	124		
Db	261	CTGC	264		

RESULT 5			
BD096983			
LOCUS	754 bp	DNA	linear
DEFINITION	SAG:apoptosis sensitivity gene.		PAT 27-AUG-2002
ACCESSION	BD096989		
VERSION	BD096989.1 GI:22642577		
KEYWORDS	JP 2001526063-A/24.		
SOURCE	unidentified		
ORGANISM	unidentified		
REFERENCE	1. (bases 1 to 754)		
AUTHORS	Sun,Y.		
TITLE	SAG:apoptosis sensitivity gene		
JOURNAL	Patent: JP 2001526063-A 24 18-DEC-2001;		
COMMENT	WARNER LAMBERT CO		
	OS unidentified		
	DB TRIMMING:300 1/24		

```

FEATURES
    source
        PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI
        Y1 SUN
        PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,
        PC A61P39/06
        PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC
        C12N5/10,C12N01/58
        PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC
        Strandedness: Double;
        CC Topology: Linear;
        CC SAG:apoptosis sensitivity gene
        FH Key Location/Qualifiers
        FT CDS 1..339
        FT mat peptide 1..339.
        Location/Qualifiers
            1..754
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 206 a 155 c 201 g 192 t
ORIGIN

```

Query Match	Score	DB	Length
Best Local Similarity	100.0%		754
Matches	124	Conservative	0; Mismatches 0; Indels 0; Gaps 0;

  

Query	Score	DB	Length
1 CGATACGTGGCCCATCTGCAGAGGTCACAGGTGATGATGCTTGTCTAGATGTCAGCTGA	60		
141 CGATACGTGGCCCATCTGCAGAGGTCACAGGTGATGATGATGCTTGTAGATGTCAGCTGA	200		
61 AAAAAACAAGAGGACGTGTGTGTGGTCGTGGGAGATGTATCATCTTCCCAACTG	120		
201 AAAAAACAAGAGGACGTGTGTGTGGTCGTGGGAGATGTATCATCTTCCCAACTG	260		
QY	121	CTGC	124
Db	261	CTGC	264

  

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
6	BD096994	SAG:apoptosis sensitivity gene.	BD096994	GI:22642582	JP 2001526063-A/29.	unidentified	unclassified.
							1 (bases 1 to 754)
							Sun, Y.
							SAG:apoptosis sensitivity gene
							Patent: JP 2001526063-A 29 18-DEC-2001;
							WARNER LAMBERT CO
							OS Unidentified

  

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Sun, Y.	SAG:apoptosis sensitivity gene	Patent: JP 2001526063-A 29 18-DEC-2001;	
2	WARNER LAMBERT CO	OS Unidentified		

FEATURES		Location/Qualifiers
source	1..754	
	/organism="unidentified"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32644"	
BASE COUNT	206 a 155 c 201 g 192 t	
ORIGIN		
Query Match	100.0%;	Score 124; DB 6; Length 754;
Best Local Similarity	100.0%;	Pred. No. 6, 4e-66;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1 CGATACGTCGCCCATCTGCAGAGGTCACAGTGATGATGACCTGCTTATGATGTCAAGCTGA 60	
Db	141 CGATACGTCGCCCATCTGCAGAGGTCACAGTGATGATGATGACCTGCTTATGATGTCAAGCTGA 200	
Oy	61 AAACAAACAGAGACATCTGTGTGTGTGTGGGGGAATGTATATATCTTCCAAACTG 120	
Db	201 AAACAAACAGAGACATCTGTGTGTGTGTGGGGGAATGTATATATCTTCCAAACTG 260	
Oy	121 CTGC 124	
Db	261 CTGC 264	
RESULT 8		
LOCUS	AF092878	754 bp mRNA linear PRI 24-JUL-2001
DEFINITION	Homo sapiens zinc RING finger protein SAG mRNA, complete cds.	
ACCESSION	AF092878	
VERSION	AF092878.1	GI:4588033
KEYWORDS		
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 754)	
AUTHORS	Duan,H., Wang,Y., Aviram,M., Swaroop,M., Loo,J.A., Bian,J., Tian,Y., Mueller,T., Bisgaier,C.L. and Sun,Y.	
TITLE	SAG, a novel zinc RING finger protein that protects cells from apoptosis induced by redox agents	
	Mol. Cell. Biol. 19 (4), 3145-3155 (1999)	
JOURNAL	MEDLINE	99182502
PUBMED	10082581	
REFERENCE	2 (bases 1 to 754)	
AUTHORS	Sun,Y.	
TITLE	Alterations of SAG mRNA in human cancer cell lines: requirement for the RING finger domain for apoptosis protection	
	Carcinogenesis 20 (10), 1899-1903 (1999)	
JOURNAL	MEDLINE	99435944
PUBMED	10506102	
REFERENCE	3 (bases 1 to 754)	
AUTHORS	Swaroop,M., Bian,J., Aviram,M., Duan,H., Bisgaier,C.L., Loo,J.A. and Sun,Y.	
TITLE	Expression, purification, and biochemical characterization of SAG, a RING finger redox-sensitive protein	
	Free Radical Biol. Med. 27, 193-202 (1999)	
JOURNAL	REFERENCE	4 (bases 1 to 754)
AUTHORS	Swaroop,M., Wang,Y., Miller,P., Duan,H., Jackoe,T., Madore,S.J. and Sun,Y.	
TITLE	Yeast homolog of human SAG/ROC2/Rbx2/Hrt2 is essential for cell growth, but not for germination: chip profiling implicates its role in cell cycle regulation	
	Oncogene 19 (24), 2855-2866 (2000)	
JOURNAL	MEDLINE	203059864
PUBMED	10851089	
REFERENCE	5 (bases 1 to 754)	
AUTHORS	Duan,H., Tsvetkov,L.M., Liu,Y., Song,Y., Swaroop,M., Wen,R., Kung,H.P., Zhang,H. and Sun,Y.	
TITLE	Promotion of S-phase entry and cell growth under serum starvation by SAG/ROC2/Rbx2/Hrt2, an E3 ubiquitin ligase component: association with inhibition of p27 accumulation	
	Mol. Carcinog. 30 (1), 37-46 (2001)	
JOURNAL		

MEDLINE 21152847  
PubMed 11255262  
REFERENCE 6 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1998) Department of Molecular Biology,  
Parke-Davis, 2800 Plymouth Rd, Ann Arbor, MI 48105, USA

FEATURES  
source  
1..754  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3q22-q24"  
/cell\_line="Hela, D98/AH-2, HPRT-"  
1..342  
/function="growth promotion"  
/note="redox sensitive, metal binding; expression protects  
cells from apoptosis induced by redox compounds"  
/codon\_start=1  
/product="zinc RING finger protein SAG"  
/protein\_id="AAD25962.1"  
/db\_xref="GI:4588034"  
/translation="MADVEDGETCALASHSGSSGSGDPMFSLKKMNAVAMWSW  
VEDTCALCRVQVMDACLRQAEKQEDCVVWGECHNSFHNCMSLWVKQNRPLC  
QDDWVVRIGK"

CDS  
1..342  
/function="growth promotion"  
/note="redox sensitive, metal binding; expression protects  
cells from apoptosis induced by redox compounds"  
/codon\_start=1  
/product="zinc RING finger protein SAG"  
/protein\_id="AAD25962.1"  
/db\_xref="GI:4588034"  
/translation="MADVEDGETCALASHSGSSGSGDPMFSLKKMNAVAMWSW  
VEDTCALCRVQVMDACLRQAEKQEDCVVWGECHNSFHNCMSLWVKQNRPLC  
QDDWVVRIGK"

BASE COUNT 205 a 155 c 201 g 193 t

ORIGIN

Query Match 100.0%; Score 124; DB 9; Length 754;  
Best Local Similarity 100.0%; Pred. No. 6.4e-66;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATTCGACGGGTCAGGATGATGATGCTCTTAGATGCAAGCTGA 60  
141 CGATACGTGCGCCATTCGACGGGTCAGGATGATGATGCTCTTAGATGCAAGCTGA 200  
DB 61 AAACAACAAGAGAGCTGTGTGCTGGGGAGATGATATCATCTCTCCACAACG 120  
201 AAACAACAAGAGAGCTGTGTGCTGGGGAGATGATATCATCTCTCCACAACG 260  
QY 121 CTGC 124  
261 CTGC 264  
DB

RESULT 9  
LOCUS BC008627 816 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, ring finger protein 7, clone MGC:17274 IMAGE:4177613,  
mRNA, complete cds.  
ACCESSION BC008627  
VERSION BC008627.1 GI:14250388  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 816)  
Strausberg, R.  
Direct Submission  
Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center

Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRK Plate: 12 Row: h Column: 14  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Similarity but not  
identity to protein.

FEATURES  
source  
1..816  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="locusID:9616"  
/db\_xref="taxon:9606"  
/clone="MGC:17274 IMAGE:4177613"  
/issue\_type="Brain, anaplastic oligodendroglioma with  
1p/19q loss"  
/clone\_1lb="NCI CGAP Brn67"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
37..378  
/codon\_start=1  
/product="ring finger protein 7"  
/protein\_id="AAH08627.1"  
/db\_xref="GI:14250389"  
/translation="MADVEDGETCALASHSGSSGSGDPMFSLKKMNAVAMWSW  
VEDTCALCRVQVMDACLRQAEKQEDCVVWGECHNSFHNCMSLWVKQNRPLC  
QDDWVVRIGK"

CDS  
37..378  
/product="ring finger protein 7"  
/protein\_id="AAH08627.1"  
/db\_xref="GI:14250389"  
/translation="MADVEDGETCALASHSGSSGSGDPMFSLKKMNAVAMWSW  
VEDTCALCRVQVMDACLRQAEKQEDCVVWGECHNSFHNCMSLWVKQNRPLC  
QDDWVVRIGK"

BASE COUNT 226 a 178 c 213 g 199 t

ORIGIN

Query Match 100.0%; Score 124; DB 9; Length 816;  
Best Local Similarity 100.0%; Pred. No. 6.5e-66;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATTCGACGGGTCAGGATGATGATGCTCTTAGATGCAAGCTGA 60  
177 CGATACGTGCGCCATTCGACGGGTCAGGATGATGATGCTCTTAGATGCAAGCTGA 236  
DB 61 AAACAACAAGAGAGCTGTGTGCTGGGGAGATGATATCATCTCTCCACAACG 120  
237 AAACAACAAGAGAGCTGTGTGCTGGGGAGATGATATCATCTCTCCACAACG 296  
QY 121 CTGC 124  
297 CTGC 300  
DB

RESULT 10  
LOCUS BC005966 822 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, ring finger protein 7, clone MGC:14618 IMAGE:4069078,  
mRNA, complete cds.  
ACCESSION BC005966  
VERSION BC005966.1 GI:13543635  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 822)  
Strausberg, R.  
Direct Submission  
Submitted (02-APR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
Email: gcgabs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINU)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-sngc.stanford.edu>  
Contact: (Dickson, Mark) mcdpaxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINU at: <http://image.lnl.gov>  
Series: IRAL Plate: 21 Row: 0 Column: 13  
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, similarity but not identity to protein.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 822  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/db\_xref="taxon:9606"  
/clone="MGC:14618 IMAGE:4069078"  
/issue\_type="Kidney, hypernephroma"  
/clone\_lib="NIH MGC\_58"  
/lab\_host="DH105"  
/notes="Vector: pDNR-LIB"  
27.368  
/codon\_start=1  
/product="ring finger protein 7"  
/protein\_id="AAH05966.1"  
/db\_xref="GI:13543636"  
/translation="MADVDEEETCALASHSSGSGSGGDMFSLKKNNAVAMSWD  
VECDTCAICRYVMDACLRCAENKQEDCVVWGECHNSFHNCMSLWVKONRCPIC  
QDDWVVRIGK"

## CDS

BASE COUNT 240 a 175 c 210 g 197 t  
ORIGIN

Query Match 100.0%; Score 124; DB 9; Length 822;  
Best Local Similarity 100.0%; Pred. No. 6.5e-66;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGCTCCAGGTGATGATGCTGCTTAAATGTCAGACTGA 60  
|||  
Db 167 CGATACGTGCGCCATCTGCAGGCTCCAGGTGATGATGCTGCTTAAATGTCAGACTGA 226  
QY 61 AAAACAACAAGAGAGCTGTGTGTGCTGTGGGAGAAATGTAATCATTCCTTCCACACTG 120  
|||  
Db 227 AAAACAACAAGAGAGCTGTGTGTGCTGTGGGAGAAATGTAATCATTCCTTCCACACTG 286  
QY 121 CTGC 124  
|||  
Db 287 CTGC 290

RESULT 11  
AF164679 836 bp mRNA linear PRI 14-OCT-1999  
LOCUS Homo sapiens ring finger protein CKBP1 mRNA, complete cds.  
DEFINITION AF164679  
ACCESSION AF164679.1 GI:5917673  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Son,M.Y., Park,J.W., Kim,Y.S., Kang,S.W., Marshak,D.R., Park,W. and  
Bae,Y.S.  
TITLE  
Protein kinase CKII interacts with and phosphorylates the SAG  
protein containing ring-H2 finger motif

JOURNAL Biochem. Biophys. Res. Commun. 263 (3), 743-748 (1999)  
MEDLINE 99443734  
PUBMED 10512750  
REFERENCE 2 (bases 1 to 836)  
AUTHORS Son,M.-Y., Park,J.-W., Kim,Y.-S., Kang,S.-W., Marshak,D.R., Park,W. and Bae,Y.-S.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUL-1999) Department of Biochemistry, Kyungpook National University, Taegu 702-701, Korea  
Location/Qualifiers

## FEATURES

## SOURCE

1. 836  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
54.395  
/note="SAG: contains an H2 ring finger motif"  
/codon\_start=1  
/product="ring finger protein CKBP1"  
/protein\_id="AAD55984.1"  
/db\_xref="GI:5917674"  
/translation="MADVDEEETCALASHSSGSGSGGDMFSLKKNNAVAMSWD  
VECDTCAICRYVMDACLRCAENKQEDCVVWGECHNSFHNCMSLWVKONRCPIC  
QDDWVVRIGK"

## CDS

BASE COUNT 228 a 185 c 223 g 200 t  
ORIGIN

Query Match 100.0%; Score 124; DB 9; Length 836;  
Best Local Similarity 100.0%; Pred. No. 6.5e-66;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATACGTGCGCCATCTGCAGGCTCCAGGTGATGATGCTGCTTAAATGTCAGACTGA 60  
|||  
Db 194 CGATACGTGCGCCATCTGCAGGCTCCAGGTGATGATGCTGCTTAAATGTCAGACTGA 253  
QY 61 AAAACAACAAGAGAGCTGTGTGTGCTGTGGGAGAAATGTAATCATTCCTTCCACACTG 120  
|||  
Db 254 AAAACAACAAGAGAGCTGTGTGTGCTGTGGGAGAAATGTAATCATTCCTTCCACACTG 313  
QY 121 CTGC 124  
|||  
Db 314 CTGC 317

RESULT 12  
BD096988 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096988  
ACCESSION BD096988.1 GI:22642576  
VERSION JP 2001526063-A/23.  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
unclassified.  
1 (bases 1 to 754)  
REFERENCE Sun,Y.  
AUTHORS SAG:apoptosis sensitivity gene  
TITLE Patent: JP 2001526063-A 23 18-DEC-2001;  
JOURNAL WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/23  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179, 11-SRP-1998 US 60/099840, PI  
YI SUN  
PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,  
PC A61P39/06,  
PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC  
C12N5/10, C12O1/68,  
PC G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers

[illegible]

Db	149	GGCCGATCTGAGGGGTCCAGGTATGATGTCCTCTTTAGATGTCAGCTGAAAAACAAC	208
Qy	69	AAGAGACGTGTTGTGTCCTGGGAGATGTAATCAATTCCTTCACAACTGCTGC	124
Db	209	AAGAGACTGTGTTGTGTCCTGGGAGATGTAATCAATTCCTTCACAACTGCTGC	264
RESULT 14			
BD096987			
LOCUS	BD096987	754 bp	DNA
DEFINITION	SAG:apoptosis sensitivity gene.	linear	PAT 27-AUG-2002
ACCESSION	BD096987		
VERSION	BD096987.1	GI:22642575	
KEYWORDS	JP 2001526063-A/22.		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 754)		
AUTHORS	Sun,Y.		
TITLE	SAG:apoptosis sensitivity gene		
JOURNAL	Patent: JP 2001526063-A 22 18-DEC-2001; WARNER LAMBERT CO		
COMMENT	OS Unidentified PN JP 2001526063-A/22 PD 18-DEC-2001 PE 15-DEC-1998 JP 20005235451 PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI YI SUN PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00, PC A6139/06, PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC PC C12N5/10,C12Q1/68, PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC Strandedness: Double; CC Topology: linear; CC SAG:apoptosis sensitivity gene FH Key Location/Qualifiers FT CDS 1..339. FT mat_peptide 1..754 Location/Qualifiers 1..754 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"		
BASE COUNT	206 a 153 c 202 g 193 t		
ORIGIN			
Query Match	90.3%;	Score 112;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 2.1e-58;	Length 754;
Matches 112;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps 0;			
Qy	1	CGATACGTGGCCATCTGCAGGGCCGAGGATGATGATGCTGTCCTTAGATGTAAGCTGA	60
Db	141	CGATACGTGGCCATCTGCAGGGCCGAGGATGATGATGCTGTCCTTAGATGTAAGCTGA	200
Qy	61	AAACAAACAAGAGACTGTGTTGTGTCCTGGGAGATGTAATCATTCCTTC	112
Db	201	AAACAAACAAGAGACTGTGTTGTGTCCTGGGAGATGTAATCATTCCTTC	252
RESULT 15			
BD096982			
LOCUS	BD096982	754 bp	DNA
DEFINITION	SAG:apoptosis sensitivity gene.	linear	PAT 27-AUG-2002
ACCESSION	BD096982		
VERSION	BD096982.1	GI:22642570	
KEYWORDS	JP 2001526063-A/17.		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 754)		
AUTHORS	Sun,Y.		
TITLE	SAG:apoptosis sensitivity gene		

JOURNAL Patent: JP 2001526063-A 17 18-DEC-2001;  
 COMMENT WARNER LAMBERT CO  
 OS Unidentified  
 PN JP 2001526063-A/17  
 PD 18-DEC-2001  
 PF 15-DEC-1998 JP 2000525451  
 PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
 YI SUN  
 PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
 PC A61P39/06,  
 PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
 ,C12N5/10,C12O1/68,  
 PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
 Strandedness: Double;  
 CC Topology: Linear;  
 CC SAG:apoptosis sensitivity gene  
 FH Key Location/Qualifiers  
 FT CDS 1..339  
 FT mat\_peptide 1..339.  
 Location/Qualifiers  
 1..754  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t

ORIGIN

Query Match 86.3%; Score 107; DB 6; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-55;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCAGGTGATGATGCTGCTTATGATGTAAGCTGAAACAAACAGAGACT 77  
 DB 158 GCAGGGTCAGGTGATGATGCTGCTTATGATGTAAGCTGAAACAAACAGAGACT 217

QY 78 GTGTTGTGCTGGGAGAAATGTAATCATCTCTCCCAACTGCTGC 124  
 DB 218 GTGTTGTGCTGGGAGAAATGTAATCATCTCTCCCAACTGCTGC 264

RESULT 16  
 LOCUS BD096983 754 bp DNA linear PAT 27-AUG-2002  
 DEFINITION SAG:apoptosis sensitivity gene.  
 ACCESSION BD096983.1 GI:22642571  
 VERSION BD096983.1  
 KEYWORDS JP 2001526063-A/18.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 754)  
 Sun,Y.  
 SAG:apoptosis sensitivity gene  
 Patent: JP 2001526063-A 18 18-DEC-2001;  
 WARNER LAMBERT CO  
 OS Unidentified  
 PN JP 2001526063-A/18  
 PD 18-DEC-2001  
 PF 15-DEC-1998 JP 2000525451  
 PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
 YI SUN  
 PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
 PC A61P39/06,  
 PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
 ,C12N5/10,C12O1/68,  
 PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
 Strandedness: Double;  
 CC Topology: Linear;  
 CC SAG:apoptosis sensitivity gene  
 FH Key Location/Qualifiers  
 FT CDS 1..339  
 FT mat\_peptide 1..339.  
 Location/Qualifiers

FEATURES

source 1..754  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

BASE COUNT 207 a 155 c 201 g 191 t

ORIGIN

Query Match 86.3%; Score 107; DB 6; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-55;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GCAGGGTCAGGTGATGATGCTGCTTATGATGTAAGCTGAAACAAACAGAGACT 77  
 DB 158 GCAGGGTCAGGTGATGATGCTGCTTATGATGTAAGCTGAAACAAACAGAGACT 217

QY 78 GTGTTGTGCTGGGAGAAATGTAATCATCTCTCCCAACTGCTGC 124  
 DB 218 GTGTTGTGCTGGGAGAAATGTAATCATCTCTCCCAACTGCTGC 264

RESULT 17  
 LOCUS BD096986 754 bp DNA linear PAT 27-AUG-2002  
 DEFINITION SAG:apoptosis sensitivity gene.  
 ACCESSION BD096986.1 GI:22642574  
 VERSION BD096986.1  
 KEYWORDS JP 2001526063-A/21.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 754)  
 Sun,Y.  
 SAG:apoptosis sensitivity gene  
 Patent: JP 2001526063-A 21 18-DEC-2001;  
 WARNER LAMBERT CO  
 OS Unidentified  
 PN JP 2001526063-A/21  
 PD 18-DEC-2001  
 PF 15-DEC-1998 JP 2000525451  
 PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
 YI SUN  
 PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
 PC A61P39/06,  
 PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
 ,C12N5/10,C12O1/68,  
 PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
 Strandedness: Double;  
 CC Topology: Linear;  
 CC SAG:apoptosis sensitivity gene  
 FH Key Location/Qualifiers  
 FT CDS 1..339  
 FT mat\_peptide 1..339.  
 Location/Qualifiers

FEATURES

source 1..754  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

BASE COUNT 207 a 154 c 201 g 192 t

ORIGIN

Query Match 83.1%; Score 103; DB 6; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 9e-53;  
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCGGCGGCATCTGACAGGCTGACAGTGTGATGCTGTCTTATGATGTAAGCTGA 60  
 DB 141 CGATCGGCGGCATCTGACAGGCTGACAGTGTGATGCTGTCTTATGATGTAAGCTGA 200

QY 61 AAACAACAAGAGACTGTGTTGTGCTGGGAGAAATGTAAT 103  
 DB 201 AAACAACAAGAGACTGTGTTGTGCTGGGAGAAATGTAAT 243

RESULT 18  
BD096990 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096990  
ACCESSION BD096990.1 GI:22642578  
VERSION JP 2001526063-A/25.  
KEYWORDS unclassified  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 25 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/25  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
BASE COUNT 208 a 152 c 202 g 192 t  
Query Match 83.1%; Score 103; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 9e-53;  
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATACGTGGCCATCTGCGAGGCTCCAGCGATGATGCTGCTTAGATGTAACCTGA 60  
DB 141 CGATACGTGGCCATCTGCGAGGCTCCAGCGATGATGCTGCTTAGATGTAACCTGA 200  
QY 61 AAACAACAAGAGAGACTGTGTGCTGCTGCGAGATGTAAT 103  
DB 201 AAACAACAAGAGAGACTGTGTGCTGCGAGATGTAAT 243  
RESULT 19  
BD096985 754 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096985  
ACCESSION BD096985.1 GI:22642573  
VERSION JP 2001526063-A/20.  
KEYWORDS unclassified  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 20 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/20  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI

YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339  
FT Location/Qualifiers  
FEATURES  
source 1..754  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
BASE COUNT 206 a 155 c 201 g 192 t  
Query Match 78.2%; Score 97; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 5.1e-49;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATACGTGGCCATCTGCGAGGCTCCAGCGATGATGCTGCTTAGATGTAACCTGA 60  
DB 141 CGATACGTGGCCATCTGCGAGGCTCCAGCGATGATGCTGCTTAGATGTAACCTGA 200  
QY 61 AAACAACAAGAGAGACTGTGTGCTGCTGCGAGAGA 97  
DB 201 AAACAACAAGAGAGACTGTGTGCTGCTGCGAGAGA 237  
RESULT 20  
BD096973 747 bp DNA linear PAT 27-AUG-2002  
LOCUS SAG:apoptosis sensitivity gene.  
DEFINITION BD096973  
ACCESSION BD096973.1 GI:22642561  
VERSION JP 2001526063-A/8.  
KEYWORDS unclassified  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 747)  
AUTHORS Sun, Y.  
TITLE SAG:apoptosis sensitivity gene  
JOURNAL Patent: JP 2001526063-A 8 18-DEC-2001;  
WARNER LAMBERT CO  
COMMENT OS Unidentified  
PN JP 2001526063-A/8  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C12Q1/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..270  
FT mat\_peptide 1..270  
FT Location/Qualifiers  
FEATURES  
source 1..747  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
BASE COUNT 204 a 155 c 197 g 191 t

Query Match 72.6%; Score 90; DB 6; Length 747;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-44;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GATGCTGCTTGTAGATGTCAGCTGAAACAAACAGAGAGACTGTGTGTGCTGGGGA 94  
 |||||  
 DB 168 GATGCTGCTTGTAGATGTCAGCTGAAACAAACAGAGAGACTGTGTGTGCTGGGGA 227  
 |||||

QY 95 GAATGTAATCATCTCTTCCACAACTGCTGC 124  
 |||||  
 DB 228 GAATGTAATCATCTCTTCCACAACTGCTGC 257  
 |||||

RESULT 21  
 AF312226 674 bp mRNA linear PRI 06-SEP-2001  
 LOCUS Homo sapiens SAG splice variant mRNA, complete cds.  
 DEFINITION AF312226  
 ACCESSION AF312226.1 GI:13649605  
 VERSION AF312226.1 GI:13649605  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 674)  
 Duan, H., Wang, Y., Aviram, M., Swaroop, M., Loo, J. A., Bian, J.,  
 Tian, Y., Mueller, T., Bieganski, C. L., and Sun, Y.,  
 SAG, a novel zinc RING finger protein that protects cells from  
 apoptosis induced by redox agents  
 Mol. Cell. Biol. 19 (4), 3145-3155 (1999)  
 99182502  
 10082581

REFERENCE 2 (bases 1 to 674)  
 Swaroop, M., Gosink, M., and Sun, Y.,  
 SAG/ROC2/Rbx2/Hrt2, a component of SCP E3 ubiquitin ligase: genomic  
 structure, a splicing variant, and two family pseudogenes  
 DNA Cell Biol. 20 (7), 425-434 (2001)  
 21398045  
 11506706

REFERENCE 3 (bases 1 to 674)  
 Sun, Y.  
 Direct Submission  
 Submitted (09-OCT-2000) Molecular Biology, Pfizer Global Research  
 and Development, 2800 Plymouth Road, Ann Arbor, MI 48105, USA  
 JOURNAL  
 TITLE  
 AUTHORS  
 PUBLISHED  
 MEDLINE  
 PUBMED

FEATURES  
 source  
 1..674  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3q22-q24"  
 1..327  
 /note="SAG-V; lacks zinc RING finger domain; has no  
 activity as a component of SCF E3 ubiquitin ligase"  
 /codon\_start=1  
 /product="SAG splice variant"  
 /protein\_id="AAK37450.1"  
 /db\_xref="GI:13649606"  
 /translation="MADVEDGETCALASHSGSSGSKGDMFSLIKKMAVAMWSW  
 VEDTCAICRVQVDEIGIVRWSEALNLIINSEMGFDCRSSTALAVPSVLSHSP  
 CLDDHR"

BASE COUNT 163 a 154 c 194 g 163 t  
 ORIGIN

Query Match 71.8%; Score 89; DB 9; Length 674;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-44;  
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ATGCTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 95  
 |||||  
 DB 440 ATGCTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 499  
 |||||

QY 96 AATGTAATCATCTCTTCCACAACTGCTGC 124  
 |||||

DB 500 AATGTAATCATCTCTTCCACAACTGCTGC 528  
 |||||

RESULT 22  
 BD096984 754 bp DNA linear PAT 27-AUG-2002  
 LOCUS BD096984  
 DEFINITION SAG:apoptosis sensitivity gene.  
 ACCESSION BD096984  
 VERSION BD096984.1 GI:22642572  
 KEYWORDS JP 2001526063-A/19.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 754)  
 Sun, Y.  
 SAG:apoptosis sensitivity gene  
 Patent: JP 2001526063-A 19-18-DEC-2001;  
 WARNER LAMBERT CO  
 COMMENT OS Unidentified

OS Unidentified  
 PN JP 2001526063-A/19  
 PD 18-DEC-2001  
 PF 15-DEC-1998 JP 2000525451  
 PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI  
 YI SUN

PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,  
 PC A61P39/06,  
 PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC  
 , C12N5/10, C12Q1/68,  
 PC G01N33/50, G01N33/68, C12N15/00, A61K37/02, C12N5/00 CC  
 Strandedness: Double;  
 CC Topology: linear;  
 CC SAG:apoptosis sensitivity gene  
 FH Key Location/Qualifiers  
 FT CDS 1..339  
 FT mat\_peptide 1..339.  
 FT Location/Qualifiers

FEATURES  
 source  
 1..754  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t  
 ORIGIN

Query Match 66.9%; Score 83; DB 6; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 3e-40;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 GTCTTAGATGTCAGCTGAAACAAACAGAGAGACTGTGTGTGCTGGGAGATGTA 101  
 |||||  
 DB 182 GTCTTAGATGTCAGCTGAAACAAACAGAGAGACTGTGTGTGCTGGGAGATGTA 241  
 |||||

QY 102 ATCATTCCTTCCACAACTGCTGC 124  
 |||||  
 DB 242 ATCATTCCTTCCACAACTGCTGC 264  
 |||||

RESULT 23  
 BD096992 754 bp DNA linear PAT 27-AUG-2002  
 LOCUS BD096992  
 DEFINITION SAG:apoptosis sensitivity gene.  
 ACCESSION BD096992  
 VERSION BD096992.1 GI:22642580  
 KEYWORDS JP 2001526063-A/27.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 754)  
 Sun, Y.  
 SAG:apoptosis sensitivity gene  
 Patent: JP 2001526063-A 27-18-DEC-2001;  
 WARNER LAMBERT CO  
 COMMENT OS Unidentified



PN JP 2001526063-A/27  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 200525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C1201/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FEATURES  
source Location/Qualifiers  
1..754  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 206 a 155 c 201 g 192 t  
ORIGIN  
Query Match 61.3%; Score 76; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 7,2e-36;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCATACGTCCGCATCTGCAGGCTCCAGGTGATGCTGCTTTAGATGTCAGTCA 60  
141 CCATACGTCCGCATCTGCAGGCTCCAGGTGATGCTGCTTTAGATGTCAGTCA 200  
Db 61 AAACAACAAGAGGAC 76  
201 AAACAACAAGAGGAC 216  
QY  
Db  
RESULT 24  
BD096991 754 bp DNA linear PAT 27-AUG-2002  
LOCUS  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096991.1 GI:22642579  
VERSION BD096991.1  
KEYWORDS JP 2001526063-A/26.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 754)  
Sun,Y.  
SAG:apoptosis sensitivity gene  
TITLE Patient: JP 2001526063-A 26 18-DEC-2001;  
JOURNAL WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/26  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 200525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C1201/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FEATURES  
source Location/Qualifiers  
1..754  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 206 a 155 c 201 g 192 t  
ORIGIN  
Query Match 59.7%; Score 74; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1,3e-34;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 51 GTCAGCTGAACAAACAAGAGAGCTGTGCTGCTGGGAGAAATGATCATTCCT 110  
191 GTCAGCTGAACAAACAAGAGAGCTGTGCTGCTGGGAGAAATGATCATTCCT 250  
Db 111 TCACAACTGCTGC 124  
251 TCACAACTGCTGC 264  
QY  
Db  
RESULT 25  
BD096993 754 bp DNA linear PAT 27-AUG-2002  
LOCUS  
DEFINITION SAG:apoptosis sensitivity gene.  
ACCESSION BD096993.1 GI:22642581  
VERSION BD096993.1  
KEYWORDS JP 2001526063-A/28.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 754)  
Sun,Y.  
SAG:apoptosis sensitivity gene  
TITLE Patient: JP 2001526063-A 28 18-DEC-2001;  
JOURNAL WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/28  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 200525451  
PR 19-DEC-1997 US 60/068179,11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
,C12N5/10,C1201/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC SAG:apoptosis sensitivity gene  
FH Key Location/Qualifiers  
FT CDS 1..339  
FT mat\_peptide 1..339.  
FEATURES  
source Location/Qualifiers  
1..754  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 207 a 155 c 201 g 191 t  
ORIGIN  
Query Match 59.7%; Score 74; DB 6; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1,3e-34;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 51 GTCAGCTGAACAAACAAGAGAGCTGTGCTGCTGGGAGAAATGATCATTCCT 110  
191 GTCAGCTGAACAAACAAGAGAGCTGTGCTGCTGGGAGAAATGATCATTCCT 250  
Db 111 TCACAACTGCTGC 124  
251 TCACAACTGCTGC 264  
QY  
Db  
RESULT 26  
AC112771 35638 bp DNA linear PRI 28-SEP-2002  
LOCUS

DEFINITION	Homio sapiens 3 BAC RP11-144C9 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION	AC112771
VERSION	AC112771.6
KEYWORDS	GI:22003982
SOURCE	HTG
ORGANISM	Homio sapiens (human)
REFERENCE	1 (bases 1 to 35638)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oniola,B., All-oman,F.R., Allen,C., Alshrocks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,U., Benton,J., Binaige,K.C., Blankenburg,K., Bonini,D., Bouck,U., Bowe,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Bulay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.C., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huij,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlovic,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W., Lousaged,H., Lozard,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Nagua,P., Martin,R., Martindale,A., Martinez,M., Massey,E., Mawlinthey,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Moysan,M., Moorishi,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Oghu,M., Okwom,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,J., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojtas,A., RojudoKhan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitarli,N., Sisson,I., Sodergren,E., Sonalik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umami,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 35638)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (24-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 35638)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 35638)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	5 (bases 1 to 35638)
AUTHORS	Worley,K.C.

**TITLE**  
**JOURNAL**  
 Direct Submission  
 Submitted (28-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 30, 2002 this sequence version replaced gi:20976460.  
**COMMENT**  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)  
 CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.  
**ANNOTATION OF FEATURES:**  
 STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarly (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.  
 SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.  
 QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:  
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1. 35638  
   /organism="Homo sapiens"  
   /mol\_type="genomic DNA"  
   /db\_xref="taxon:9606"  
   /chromosome="3"  
   /clone="RP11-144C9"  
   1. 1895  
   /note="overlaps bases 56107..58001 of clone AC092978"  
   /function="Clone overlap"  
   14. 341  
   /strand.name="SHGC-105132"  
   complement(1283..1580)  
   /rpt\_family="AluSx"  
   1618..1906  
   /rpt\_family="AluSx"  
   1907..1986  
   /rpt\_family="(GGGGA)n"  
   complement(2109..2203)  
   /rpt\_family="L1M4"  
   complement(2204..2463)  
   /rpt\_family="AluSx"  
   complement(2464..2700)  
   /rpt\_family="L1M4"  
   complement(2701..2819)  
   /rpt\_family="AluSc"  
   complement(2843..3539)  
   /rpt\_family="L1M4"  
   complement(3541..3642)  
   /rpt\_family="X"  
   3643..3758

```

repeat_region      /rpt_family="FLAM_A"
                    complement(3777..4993)
                    /rpt_family="L1MEC"
repeat_region      /rpt_family="Alusq"
                    complement(4994..5287)
repeat_region      /rpt_family="Alusq"
                    complement(5288..5404)
repeat_region      /rpt_family="L1MEC"
                    complement(5405..5716)
repeat_region      /rpt_family="Alusx"
                    complement(5717..6153)
repeat_region      /rpt_family="L1MEC"
                    complement(6154..6379)
repeat_region      /rpt_family="Alusq/x"
                    complement(6380..6394)
repeat_region      /rpt_family="L1MEC"
                    complement(6427..6570)
repeat_region      /rpt_family="MER91A"
                    complement(7418..7561)
repeat_region      /rpt_family="Alusq/x"
                    complement(8094..8139)
repeat_region      /rpt_family="TG)n"
                    complement(10186..10802)
repeat_region      /rpt_family="L2"
                    complement(11949..12489)
repeat_region      /rpt_family="HAL1"
                    complement(12490..12772)
repeat_region      /rpt_family="Alusx"
                    complement(12773..12804)
repeat_region      /rpt_family="(TAAA)n"
                    complement(12805..12820)
repeat_region      /rpt_family="Alusx"
                    complement(12821..12896)
repeat_region      /rpt_family="HAL1"
                    complement(12989..13063)
repeat_region      /rpt_family="Alusq/x"
                    complement(13069..13452)
repeat_region      /rpt_family="HAL1"
                    complement(13453..13585)
repeat_region      /rpt_family="FLAM_C"
                    complement(13586..13609)
repeat_region      /rpt_family="HAL1"
                    complement(13619..13654)
repeat_region      /rpt_family="G-rich"
                    complement(13653..13809)
repeat_region      /rpt_family="GA-rich"
                    complement(13819..14048)
repeat_region      /rpt_family="AluJo"
                    complement(14077..14170)
repeat_region      /rpt_family="MER113"
                    complement(14655..14830)
repeat_region      /rpt_family="MUT1J"
                    complement(14852..15139)
repeat_region      /rpt_family="Alusp"

```

Query Match 39.5%; Score 49; DB 9; Length 35638;  
 Best Local Similarity 100.0%; Pred. No. 74e-19;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 GATGCTGTTAGATGTCAGCTGAAACAAACAGAGACTGTGTTG 83  
 Db 30671 GATGCTGTTAGATGTCAGCTGAAACAAACAGAGACTGTGTTG 30719

RESULT 27  
 AC068693/c 125041 bp DNA 1linear HTG 04-MAY-2001  
 LOCUS Homo sapiens chromosome 17 clone RP11-144C9 map 17, WORKING DRAFT  
 DEFINITION  
 SEQUENCE, 25 unordered pieces.  
 AC068693  
 ACCESSION  
 VERSION AC068693.2 GI:8139716  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

# REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 125041)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 17, clone RP11-144C9  
 Unpublished  
 2 (bases 1 to 125041)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Baeten, V., Beda, F.,  
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heatord, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehotzky, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,  
 McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R.,  
 Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Subdivision  
 Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 1, 2000 this sequence version replaced gi:7712215.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: 144 C 9  
 Center clone name: 144 C 9  
 Summary Statistics  
 Sequencing vector: M13; W77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 112098 bases at least Q40  
 Consensus quality: 118849 bases at least Q30  
 Consensus quality: 121119 bases at least Q20  
 Insert size: 158000; agarose-fp  
 Insert size: 122641; sum-of-contigs  
 Quality coverage: 3.8 in Q20 bases; agarose-fp  
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 25 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 1019: contig of 1019 bp in length  
 \* 1020 1119: gap of 100 bp  
 \* 1120 3013: contig of 1894 bp in length  
 \* 3014 3113: gap of 100 bp  
 \* 3114 5008: contig of 1895 bp in length  
 \* 5009 5108: gap of 100 bp  
 \* 5109 7154: contig of 2046 bp in length  
 \* 7155 7254: gap of 100 bp  
 \* 7255 10405: contig of 3151 bp in length  
 \* 10406 10505: gap of 100 bp

\* 10506 12914: contig of 2409 bp in length  
\* 12915 13014: gap of 100 bp  
\* 13015 13781: contig of 767 bp in length  
\* 13782 13881: gap of 100 bp  
\* 13882 17746: contig of 3865 bp in length  
\* 17747 17846: gap of 100 bp  
\* 17847 20494: contig of 2648 bp in length  
\* 20495 20594: gap of 100 bp  
\* 20595 22173: contig of 1579 bp in length  
\* 22174 22273: gap of 100 bp  
\* 22274 25788: contig of 3515 bp in length  
\* 25789 30218: gap of 100 bp  
\* 30219 30318: gap of 100 bp  
\* 30319 33343: contig of 3025 bp in length  
\* 33344 33443: gap of 100 bp  
\* 33444 36428: contig of 2985 bp in length  
\* 36429 36528: gap of 100 bp  
\* 36529 40028: contig of 3500 bp in length  
\* 40029 40128: gap of 100 bp  
\* 40129 45543: contig of 5415 bp in length  
\* 45544 45643: gap of 100 bp  
\* 45644 52370: contig of 6727 bp in length  
\* 52371 52470: gap of 100 bp  
\* 52471 56859: contig of 4399 bp in length  
\* 56860 62452: gap of 100 bp  
\* 62453 62552: gap of 100 bp  
\* 62553 70768: contig of 8216 bp in length  
\* 70769 70868: gap of 100 bp  
\* 70869 76175: contig of 5307 bp in length  
\* 76176 76275: gap of 100 bp  
\* 76276 84628: contig of 8353 bp in length  
\* 84629 84729: gap of 100 bp  
\* 84729 95400: contig of 10672 bp in length  
\* 95401 95500: gap of 100 bp  
\* 95501 111561: contig of 16061 bp in length  
\* 111562 111662: gap of 100 bp  
\* 111662 125041: contig of 13380 bp in length.

## FEATURES

source

1.125041 /organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-14C9"  
/clone\_lib="RPC1-11 Human Male BAC"  
1.1019 /note="assembly\_fragment"  
1120.3013 /note="assembly\_fragment"  
3114.5008 /note="assembly\_fragment"  
5109.7154 /note="assembly\_fragment"  
7255.10405 /note="assembly\_fragment"  
10506.12914 /note="assembly\_fragment"  
13015.13781 /note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
13882.17746 /note="assembly\_fragment"  
17847.20494 /note="assembly\_fragment"  
20595.22173 /note="assembly\_fragment"  
clone\_end:r7  
vector\_side:right"  
22274.25788

misc\_feature /note="assembly\_fragment"  
25889..30218 /note="assembly\_fragment"  
misc\_feature 30319..33343 /note="assembly\_fragment"  
misc\_feature 33444..36428 /note="assembly\_fragment"  
misc\_feature 36529..40028 /note="assembly\_fragment"  
misc\_feature 40129..45543 /note="assembly\_fragment"  
misc\_feature 45644..52370 /note="assembly\_fragment"  
misc\_feature 52471..56869 /note="assembly\_fragment"  
misc\_feature 56970..62452 /note="assembly\_fragment"  
misc\_feature 62553..70768 /note="assembly\_fragment"  
misc\_feature 70869..76175 /note="assembly\_fragment"  
misc\_feature 76276..84628 /note="assembly\_fragment"  
misc\_feature 84729..95400 /note="assembly\_fragment"  
misc\_feature 95501..111561 /note="assembly\_fragment"  
misc\_feature 111662..125041 /note="assembly\_fragment"

BASE COUNT 34737 a 26838 c 26914 g 34145 t 2407 others  
ORIGIN

Query Match 39.5%; Score 49; DB 2; Length 125041;  
Best Local Similarity 100.0%; Pred.No.8.1e-19;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GATCCCTGCTTAAATGTCAGCTGAACAAACAAAGAGACTGTGTTG 83

Db 17657 GATCCCTGCTTAAATGTCAGCTGAACAAACAAAGAGACTGTGTTG 17609

RESULT 28  
LOCUS AC106176/c 238330 bp DNA linear HTG 13-MAY-2003  
DEFINITION Rattus norvegicus clone CH230-20P16, WORKING DRAFT SEQUENCE, 3  
unordered pieces.  
ACCESSION AC106176  
VERSION AC106176.5 GI:30578781  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 238330)  
AUTHORS Muzny,D.Marie, Metzker,M.Lee, Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,B., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,  
Cardenas,V., Carter,K., Cavazos,I., Cessat,H., Center,A.,  
Chacko,U., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregiorgis,E., Geer,K., Gill,R., Gilly,M., Guerra,M., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

TITLE	JOURNAL	REFERENCE	AUTHORS	JOURNAL	COMMENT
2 (bases 1 to 238330)	Worley, K.C.	Submitted (12-JAN-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	3 (bases 1 to 238330)	Rat Genome Sequencing Consortium.
Direct Submission	Submitted (13-MAY-2003)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	On May 13, 2003 this sequence version replaced gi:24819215.		The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-----	Genome Center				
Center: Baylor College of Medicine					
Center code: BCM					
Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>					
Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a>					
-----	Project Information				
Center project name: GJOH					
Center clone name: CH230-20P16					
-----	Summary Statistics				
Assembly program: Atlas 3.0;					
Consensus quality: 229461 bases at least Q40					
Consensus quality: 231714 bases at least Q30					
Consensus quality: 233148 bases at least Q20					
Estimated insert size: 239075; sum-of-contigs estimation					
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation					

FEATURES	source
* NOTE: Estimated insert size may differ from sequence length	
* (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ).	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 3 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 236056: contig of 236056 bp in length	
* 236057 236156: gap of unknown length	
* 236157 237313: contig of 1057 bp in length	
* 237214 237313: gap of unknown length	
* 237314 238330: contig of 1017 bp in length.	
* Location/Qualifiers	
1. 238330	
/organism="Rattus norvegicus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10116"	
/clone="CH230-20016"	
1. 1582	
/note="wgs contig"	
64585 a 53668 c 53820 g 62326 t 3931 others	
misc_feature	
BASE COUNT	
ORIGIN	
Query Match	39.5%; Score 49; DB 2; Length 238330;
Best Local Similarity	100.0%; Pred. No. 8,5e-19;
Matches	49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY	35 GATGCGCTGTGATGTCAAGCTCAAGCAAAACAAGAGACGTGTG 83
Db	158930 GATGCGCTGTGATGTCAAGCTCAAGCAAAACAAGAGACGTGTG 158882
RESULT 29	
AC098496/c	
LOCUS	
DEFINITION	AC098496 245476 bp DNA linear HTG 10-MAY-2003
ACCESSION	Rattus norvegicus clone CH230-23C21, WORKING DRAFT SEQUENCE, 4
VERSION	unordered pieces.
KEYWORDS	AC098496
SOURCE	HTG: HTGS PHASE1; HTGS DRAFT; HTGS-FULLTOP.
ORGANISM	Rattus norvegicus (Norway rat)
REFERENCE	Rattus norvegicus
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
	1 (bases 1 to 245476)
	Mizny,D.Marie, Metzker,M.Lee, Abramson,S, Adams,C, Alder,J.,
	Allen,C, Allen,H, Aisbrooks,S, Amin,A, Anguiano,D,
	Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
	Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
	Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
	Bryan,N, Buhal,C, Burch,P, Burrell,K, Calderon,E,
	Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
	Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
	Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
	David,L,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
	Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
	Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
	Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
	Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
	Fraser,C,M, Gabis,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
	Gebregiorgis,E, Geer,K, Gill,I, Grady,M, Guerra,W, Guevara,W,
	Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
	Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
	Henderson,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M,
	Hollins,B, Howell,S, Huik,S, Hume,J, Idlebird,D, Jackson,A,
	Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
	Jarkapthy,S, Kelly,S, Khan,Z, King,L, Kover,C,
	Kowis,C, Kraft,C,L, Lebrow,H, Levan,J, Lewis,L, Li,Z, Liu,J,





```

CC id M06951
CC est
CC blastn
CC identity 96
CC region 7.64
CC id M06951
CC est
CC blastn
CC identity 97
CC region 2.179
CC id W38711
CC est
CC Von Heljne matrix
CC score 3.7
CC seq ETCALASHSGSSG/SK
CC n=a, g, c or t
FH Key Location/Qualifiers
FT misc_feature 89.222
FT misc_feature 26.90
FT misc_feature 89.222
FT misc_feature 33.90
FT misc_feature 45.222
FT sig_peptide 24.86
FT misc_feature 3.
Location/Qualifiers
1.224
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

BASE COUNT 41 a 60 c 83 g 38 t 2 others
ORIGIN

```

```

Query Match 34.7% Score 43; DB 6; Length 224;
Best Local Similarity 100.0%; Pred. No. 3e-15;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CGATACGTGCGCCATCTGCAGGATCCAGGTGATGATGCTGT 43
Db 164 CGATACGTGCGCCATCTGCAGGATCCAGGTGATGATGCTGT 206

```

```

RESULT 32
AC104983
LOCUS Homo sapiens chromosome 18 clone RP11-352C3 map 18, LOW-PASS
DEFINITION AC104983 59155 bp DNA linear HTG 22-DEC-2001
AC104983
SEQUENCE SAMPLING.
AC104983.1 GI:17977361
HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 59155)
Birken, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone RP11-352C3
Unpublished
2 (bases 1 to 59155)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heald, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McSheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V.,

```

# TITLE JOURNAL COMMENT

Raymond, C., Rella, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)  
Project Information  
Center project name: L22883  
Center clone name: 352\_C\_3

\*\*\*\*\*  
\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\*\*\*\*\*

```

1 725 824: contig of 724 bp in length
825 825 824: gap of 100 bp
1583 1582: contig of 758 bp in length
1683 1682: gap of 100 bp
2409 2409: contig of 727 bp in length
2510 2509: gap of 100 bp
3236 3235: contig of 726 bp in length
3336 3335: gap of 100 bp
4091 4090: contig of 755 bp in length
4191 4190: gap of 100 bp
4955 4955: contig of 765 bp in length
5055 5055: gap of 100 bp
5780 5780: contig of 725 bp in length
5881 5880: gap of 100 bp
6643 6643: contig of 763 bp in length
6744 6743: gap of 100 bp
7496 7496: contig of 753 bp in length
7497 7496: gap of 100 bp
7597 7596: gap of 100 bp
8348 8348: contig of 752 bp in length
8448 8448: gap of 100 bp
9217 9217: contig of 769 bp in length
9317 9317: gap of 100 bp
9318 9318: gap of 100 bp
10093 10093: contig of 776 bp in length
10194 10193: gap of 100 bp
10924 10924: contig of 731 bp in length
11024 11024: gap of 100 bp
11793 11793: contig of 769 bp in length
11794 11793: gap of 100 bp
11893 11893: gap of 100 bp
11894 11894: contig of 768 bp in length
12662 12662: gap of 100 bp
12761 12761: gap of 100 bp
13525 13525: contig of 764 bp in length
13625 13625: gap of 100 bp
14347 14347: contig of 722 bp in length
14447 14447: gap of 100 bp
14438 14438: gap of 100 bp
15172 15172: contig of 725 bp in length
15173 15173: gap of 100 bp
15272 15272: gap of 100 bp
15273 15273: contig of 735 bp in length
16107 16107: gap of 100 bp
16108 16108: gap of 100 bp
16855 16855: contig of 748 bp in length
16856 16856: gap of 100 bp

```



```

16956 17711: contig of 756 bp in length
* 17712 17811: gap of 100 bp
* 17812 18568: contig of 757 bp in length
* 18568 18668: gap of 100 bp
* 18668 19421: contig of 753 bp in length
* 19421 19521: gap of 100 bp
* 19521 20229: contig of 708 bp in length
* 20229 20330: gap of 100 bp
* 20330 21062: contig of 732 bp in length
* 21062 21161: gap of 100 bp
* 21161 21909: contig of 748 bp in length
* 21909 22009: gap of 100 bp
* 22009 22773: contig of 764 bp in length
* 22773 22874: gap of 100 bp
* 22874 23623: contig of 750 bp in length
* 23623 23723: gap of 100 bp
* 23723 24489: contig of 766 bp in length
* 24489 24589: gap of 100 bp
* 24589 25354: contig of 765 bp in length
* 25354 25454: gap of 100 bp
* 25454 26174: contig of 720 bp in length
* 26174 26274: gap of 100 bp
* 26274 27043: contig of 769 bp in length
* 27043 27143: gap of 100 bp
* 27143 27900: contig of 757 bp in length
* 27900 28000: gap of 100 bp
* 28000 28753: contig of 753 bp in length
* 28753 28853: gap of 100 bp
* 28853 29572: contig of 719 bp in length
* 29572 29672: gap of 100 bp
* 29672 30419: contig of 747 bp in length
* 30419 30519: gap of 100 bp
* 30519 31262: contig of 743 bp in length
* 31262 31362: gap of 100 bp
* 31362 32052: contig of 690 bp in length
* 32052 32152: gap of 100 bp
* 32152 32909: contig of 757 bp in length
* 32909 33009: gap of 100 bp
* 33009 33736: contig of 727 bp in length
* 33736 33836: gap of 100 bp
* 33836 34598: contig of 762 bp in length
* 34598 34698: gap of 100 bp
* 34698 35456: contig of 758 bp in length
* 35456 35556: gap of 100 bp
* 35556 36274: contig of 718 bp in length
* 36274 36374: gap of 100 bp
* 36374 37111: contig of 737 bp in length
* 37111 37211: gap of 100 bp
* 37211 37954: contig of 743 bp in length
* 37954 38054: gap of 100 bp
* 38054 38783: contig of 725 bp in length
* 38783 38883: gap of 100 bp
* 38883 39686: contig of 803 bp in length
* 39686 39786: gap of 100 bp
* 39786 40535: contig of 749 bp in length
* 40535 40635: gap of 100 bp
* 40635 41387: contig of 752 bp in length
* 41387 41487: gap of 100 bp
* 41487 42254: contig of 767 bp in length
* 42254 42354: gap of 100 bp
* 42354 43046: contig of 692 bp in length
* 43046 43146: gap of 100 bp
* 43146 43885: contig of 739 bp in length
* 43885 43985: gap of 100 bp
* 43985 44741: contig of 756 bp in length
* 44741 44841: gap of 100 bp
* 44841 45604: contig of 763 bp in length
* 45604 45704: gap of 100 bp
* 45704 46429: contig of 725 bp in length
* 46429 46529: gap of 100 bp
* 46529 47244: contig of 715 bp in length
* 47244 47344: gap of 100 bp
* 47344 48088: contig of 744 bp in length

```

```

* 48089 48188: gap of 100 bp
* 48188 48948: contig of 760 bp in length
* 48948 49048: gap of 100 bp
* 49048 49803: contig of 755 bp in length
* 49803 49903: gap of 100 bp
* 49903 50648: contig of 745 bp in length
* 50648 50748: gap of 100 bp
* 50748 51506: contig of 758 bp in length
* 51506 51606: gap of 100 bp
* 51606 52345: contig of 739 bp in length
* 52345 52445: gap of 100 bp
* 52445 53185: contig of 740 bp in length
* 53185 53285: gap of 100 bp
* 53285 54059: contig of 774 bp in length
* 54059 54159: gap of 100 bp
* 54159 54901: contig of 742 bp in length
* 54901 55001: gap of 100 bp
* 55001 55740: contig of 739 bp in length
* 55740 55840: gap of 100 bp
* 55840 56577: contig of 737 bp in length
* 56577 56677: gap of 100 bp
* 56677 57433: contig of 756 bp in length
* 57433 57533: gap of 100 bp
* 57533 58290: contig of 757 bp in length

Query Match      34.7%  Score 43; DB 2; Length 59155;
Best Local Similarity 100.0%; Pred.No. 4.4e-15;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      47 AGATGTCAGCTGAACACACAGAGAGCTGTGTGCTCT 89
Db      7777 AGATGTCAGCTGAACACACAGAGAGCTGTGTGCTCT 7819

```

```

RESULT 33
AC108040/c
LOCUS      AC108040      163521 bp      DNA      linear      HTG 30-JAN-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-210010, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION  AC108040
VERSION    AC108040.2  GI:18425316
KEYWORDS   HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULFROP.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 163521)
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 163521)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (24-JAN-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          On Jan 30, 2002 this sequence version replaced gi:18308830.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
          Project Information
Center project name: H.NH0210010
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162304 bases at least Q40

```

Consensus quality: 162444 bases at least Q30  
 Consensus quality: 162529 bases at least Q20  
 Insert size: 189000; agarose-fp  
 Insert size: 163221; sum-of-contigs  
 Quality coverage: 8.88 in Q20 bases; agarose-fp  
 Quality coverage: 8.85 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1564: contig of 1564 bp in length  
 \* 1565 1664: gap of unknown length  
 \* 1665 88984: contig of 87320 bp in length  
 \* 88985 89084: gap of unknown length  
 \* 89085 112702: contig of 23618 bp in length  
 \* 112703 112802: gap of unknown length  
 \* 112803 163521: contig of 50719 bp in length.

FEATURES  
 source 1. 163521

/organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"

misc\_feature 1. 1564  
 /note="assembly\_name:Contig25"

misc\_feature 1665. 88984  
 /note="assembly\_name:Contig28"

misc\_feature 89085. 112702  
 /note="assembly\_name:Contig26"

misc\_feature 112803. 163521  
 /note="assembly\_name:Contig27"  
 BASE COUNT 51049 a 32426 c 33346 g 46399 t 301 others  
 ORIGIN

Query Match 34.7%; Score 43; DB 2; Length 163521;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-15;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ACATGTCACAGCTGAAACAAACAGAGAGCTGTTGTCCT 89  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 80232 AGATGTCACAGCTGAAACAAACAGAGAGCTGTTGTCCT 80190

RESULT 34  
 ACO92418/c 164500 bp DNA linear PRI 18-DEC-2002  
 LOCUS Homo sapiens chromosome 3 clone RP11-229A12, complete sequence.  
 DEFINITION ACO92418 AC016923  
 AC092418.3 GI:27228861  
 VERSION HTG.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 164500)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 164500)  
 Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 164500)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Phelps, K.A., Buckley, D., Raymond, C. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 4 (bases 1 to 164500)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 COMMENT On Dec 18, 2002 this sequence version replaced gi:21166206.

Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: uwgc@u.washington.edu  
 Drafting Center: BCM

----- Project Information  
 Center project name: chr-3  
 Center clone name: RP11-229A12 (bc0315)

----- Summary Statistics  
 Sequencing vector: unknown; 49% of reads  
 Sequencing vector: plasmid; 51% of reads  
 Chemistry: Dye-terminator ET; 48% of reads  
 Chemistry: Unknown; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 51% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 164200 bases at least Q40  
 Consensus quality: 164447 bases at least Q30  
 Consensus quality: 164498 bases at least Q20  
 Insert size: 164500; sum-of-contigs  
 Quality coverage: 7.7x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:  
 5': RP11-120C2 (UWGC:bc0239) AC121250, 4576-bp overlap  
 3': RP11-755B10 (UWGC:bc0671) AC114480

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an  
 alternate chemistry or covered by high quality data (i.e., Phred  
 quality >= 30); an attempt was made to resolve all sequencing  
 problems, such as compressions and repeats; all regions were  
 covered by at least one plasmid subclone or more than one M13  
 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest  
 fingerprinting. Comparison of the experimentally derived digest  
 fragments with sequence-predicted fragments is given below.  
 The electronically-digested sequence consists of both insert and  
 vector, in order to accurately represent the entire circular BAC.  
 Small fragments below a variable cutoff (approximately 400-800 bp)  
 are not resolved in the fingerprint and hence do not appear  
 in the table. There are no significant remaining discrepancies  
 between the experimental and predicted values. Uniquely ordered  
 fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SegDerMap	FngPrnt	SegDerMap	FngPrnt	SegDerMap	FngPrnt
-----	-----	-----	-----	-----	-----
8696	8826	826	813	4194	4490
6	<800	6382	6597	2067	2161
4056	4085	512	<800	7354	7474
2246	2301	449	<800	13554	13505
4995	4994	1929	1902	1393	1424
19446	19781	11580	11448	2689	2866
2094	2106	1602	1589	3504	3735
3372	3414	2956	2966	1063	1122
1138	1111	2707	2764	5756	5729
1765	1764	27884	27982	6206	6239
4097	4085	390	<800	3403	3564
2949	2978	2569	2601	2216	2353
3174	3147	2228	2241	782	<800
1114	1111	3882	3909	4423	4633
2427	2451	1295	1271	450	<800
15	<800	202	<800	601	<800
301	<800	7894	7858	3318	3413
6455	6537	504	<800	8046	8082
118	<800	782	<800	1110	1122
4564	4499	10447	10305	4632	4633
3599	3545	533	<800	477	<800
9664	9641	7717	7858	157	<800
250	<800	678	<800	5035	5074
15160	14802	1701	1643	761	<800
2025	1971	390	<800	441	<800
1905	1898	1946	1902	5627	5729
1412	1386	101	<800	7715	7683
3186	3147	637	<800	11604	11214
6226	6297	312	<800	2582	2697
4066	4085	2950	2966	3227	3413
1785	1764	1884	1902	4021	4204
133	<800	4779	4731	2721	2866
5971	5982	5304	5321	6232	6239
3070	3147	12363	12136	4125	4204

```

-----
620 <800 282 <800 2815 3011
5882 5982 1170 1151 4140 4204
4146 4085 4004 4067 8173 8082
805 825 2339 2360 552 <800
12533 12212 5102 5148 3690 3735
5720 5773 21981 21990 14118 14343
5473 5505 10346 10305 3444 3564
117 <800 215 <800 6120 6239
639 <800 784 813
23 <800
7100 7064
-----
FEATURES
source Location/Qualifiers
1..164500
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-229A12"
/clone_lib="RP11 human BAC library 11"
137635..137636
/note="1810 and 9-bp target site duplication excised from
between bases 137635 and 138973
CGATGAAATCCCTTATGATTTTGGTAAATCATTAAGTGAATGATCATCT
TGTCAATGATCAATGTTGCGCAAAATCAATATCAGACACAAAGATGCGAA

misc_feature
34.7%: Score 43; DB 9; Length 164500;
Best Local Similarity 100.0%; Freq. NO. 4.7e-15;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 AGATGTCAGCTGTAACAAACAAGAGACTGTGTGTCT 89
|||||
Db 110158 AGATGTCAGCTGTAACAAACAAGAGACTGTGTGTCT 110116

RESULT 35
BD096974 706 bp DNA linear PAT 27-AUG-2002
BD096974
LOCUS SAG:apoptosis sensitivity gene.
DEFINITION SAG:apoptosis sensitivity gene.
ACCESSION BD096974.1 GI:22642562
VERSION BD096974.1 GI:22642562
KEYWORDS JP 2001526063-A/9.
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 706)
AUTHORS Sun, Y.
TITLE SAG:apoptosis sensitivity gene
JOURNAL Patent: JP 2001526063-A 9 18-DEC-2001;
JOURNAL WARNER LAMBERT CO
COMMENT OS Unclassified
PN JP 2001526063-A/9
PD 18-DEC-2001
PF 15-DEC-1998 JP 2000525451
PR 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI
YI SUN
PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P17/02, A61P35/00,
PC A61P39/06,
PC A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC
PC C12N5/10, C12O1/68,
PC GO1N33/50, GO1N33/68, C12N15/00, A61K37/02, C12N5/00 CC
Strandedness: Double;

```

```

CC      Topology: Linear;
CC      SAG:apoptosis sensitivity gene
FH      Key      Location/Qualifiers
FT      CDS      1..291
FT      mat_peptide 1..291.
FEATURES
  source
    1..706
    /organism="unidentified"
    /mol_type="genomic DNA"
    /db_xref="taxon:32644"
BASE COUNT      189 a      147 c      189 g      181 t
ORIGIN
Query Match      33.9%; Score 42; DB 6; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      83 GTGGCTGGGAGAAATGTAATCATTCCTCCACCACTGCTGC 124
DB      175 GTGGCTGGGAGAAATGTAATCATTCCTCCACCACTGCTGC 216

RESULT 36
AC095698/c
LOCUS      AC095698      254832 bp      DNA      linear      HTG_09-NOV-2002
DEFINITION      Rattus norvegicus clone CH230-9D13, WORKING DRAFT SEQUENCE.
ACCESSION      AC095698
VERSION      AC095698.6 GI:24817898
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 254832)
REFERENCE
  AUTHORS      Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
            Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguitano, D.,
            Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
            Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
            Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
            Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
            Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
            Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
            Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
            Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dedertch, D.,
            Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
            Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
            Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
            Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
            Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
            Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
            Guneratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
            Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
            Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
            Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
            Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
            Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,
            Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
            Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
            Lornshuber, L., Loulsegod, H., Lozado, R.J., Lu, X., Ma, D.,
            Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
            Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
            Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
            Milsavljeyvic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
            Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Natf, L.,
            Nankervys, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
            Nwackeleneh, O., Okunonu, G., Olarnungsoon, A., Pal, S., Parks, K.,
            Pasernak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C.,
            Plopper, F., Poindexter, A., Popovic, D., Pritts, B., Pu, L.-L.,
            Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,
            Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
            Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
            Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

```

```

TITLE      JOURNAL
REFERENCE
  AUTHORS      Wotley, K.C.
  TITLE      Direct Submission
  JOURNAL      Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 254832)
REFERENCE
  AUTHORS      Human Genome Sequencing Consortium.
  TITLE      Direct Submission
  JOURNAL      Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Nov 9, 2002 this sequence version replaced gi:23267899.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GDBC
            Center clone name: CH230-9D13
            ----- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 238918 bases at least Q40
            Consensus quality: 241069 bases at least Q30
            Consensus quality: 242304 bases at least Q20
            Estimated insert size: 243228; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            * 1 254832: contig of 254832 bp in length.
            Location/Qualifiers
              1..254832
              /organism="Rattus norvegicus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10116"
              /clone="CH230-9D13"
              1..1482
            misc_feature

```

/note="wgs\_end\_extension  
clone\_end:Sp6"  
2092..2921  
/note="clone\_boundary  
clone\_end:Sp6  
site:EcORI  
end\_sequence:BH339916"  
233357..235284  
/note="wgs\_contig"  
253718..254832  
/note="wgs\_contig"  
complement(253801..254228)  
/note="clone\_boundary  
clone\_end:T7  
site:EcORI  
end\_sequence:BH339912"  
BASE COUNT 63080 a 58350 c 57410 g 64047 t 11945 others  
ORIGIN

Query Match 29.8%; Score 37; DB 2; Length 254832;  
Best Local Similarity 100.0%; Pred. No. 2.8e-11;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CCATCTGCAGGTCACAGTGATGATGCTGTCTTAG 48  
|||||  
Db 146354 CCATCTGCAGGTCACAGTGATGATGCTGTCTTAG 146318

RESULT 37  
AL929565 120515 bp DNA linear ROD 16-APR-2003  
LOCUS Mouse DNA sequence from clone RP23-454N16 on chromosome 4, complete  
DEFINITION sequence.  
ACCESSION AL929565.1 GI:30014183  
VERSION AL929565.16  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 120515)  
Andrew, R.  
Direct Submission  
Submitted (16-APR-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk  
On Apr 16, 2003 this sequence version replaced gi:29825605.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.  
-----Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:

Em., ENBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-454N16 is  
from the RPCI-23 Mouse BAC library  
constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.  
-----Location/Qualifiers  
FEATURES  
source 1..120515  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-454N16"  
/clone\_lib="RPCI-23"  
BASE COUNT 34369 a 25840 c 26441 g 33865 t  
ORIGIN

Query Match 29.0%; Score 36; DB 10; Length 120515;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 GCCATCTGCAGGTCACAGTGATGATGCTGTCTT 46  
|||||  
Db 57860 GCCATCTGCAGGTCACAGTGATGATGCTGTCTT 57825

RESULT 38  
HSA323208 439 bp DNA linear PRI 18-JUL-2002  
LOCUS Homo sapiens genomic sequence surrounding Not1 site, clone  
DEFINITION NBI-073R.  
ACCESSION AJ323208  
VERSION AJ323208.1 GI:15867587  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 439)  
Kutsenko, A.S., Gazatullin, R.Z., Al-Amin, A.N., Wang, F., Khasba, S.M.,  
Podowski, R.M., Matubshkin, Y.G., Gyanchandani, A., Muravenko, O.V.,  
Levitsky, V.G., Kolchanov, N.A., Protchenov, A.I., Kashuba, V.I.,  
Kiselev, L.L., Massetman, W., Mahlestedt, C. and Zabarovsky, E.R.  
Not1 flanking sequences: a tool for gene discovery and verification  
of the human genome  
Nucleic Acids Res. 30 (14), 3163-3170 (2002)  
JOURNAL MEDLINE 22131767  
PUBMED 12136098  
REFERENCE 2 (bases 1 to 439)  
AUTHORS Zabarovsky, E.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumourbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden  
-----Location/Qualifiers  
FEATURES  
source 1..439  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="NBI-073R"  
BASE COUNT 63 a 161 c 146 g 68 t 1 others  
ORIGIN

Query Match 28.2%; Score 35; DB 9; Length 439;  
Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCATACGTCGCCCATCTGCAGGTCACAGTATG 35  
|||||  
Db 227 CCATACGTCGCCCATCTGCAGGTCACAGTATG 193

RESULT 39  
AC137264/C  
LOCUS  
DEFINITION  
AC137264 264606 bp DNA linear HTG 19-NOV-2002  
Rattus norvegicus clone CH230-66B6, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
3 unordered pieces.  
AC137264  
AC137264.1 GI:25073249  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 264606)  
REFERENCE  
AUTHORS  
Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Altschrocks,S, Amin,A, Angiano,D, Anyalebechi,V, Ayogei,A, Ayodeji,M, Baca,E, Baden,H, Balwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Catroll,L, De Anda,C, Dedetich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dim,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagge,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Gervara,M, Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,U, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huiyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jollivet,A, Karpeth,S, Kelly,S, Kelly,S, Khan,Z, King,L, Koyar,C, Kows,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensuwei,L, Loulesged,H, Lozano,R,J, Lu,X, Ma,U, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nait,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwackeleme,O, Okwuonu,G, Olarnpunsagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfamkooh,C, Plopper,F, Polindexter,A, Popovic,D, Pritmus,E, Pu,L, L, Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,U, Sanders,W, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,U, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajs,D, Speed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,U, Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabori,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasaana,D, Waldron,L, Walker,B, Wang,U, Wang,Q, Wang,S, Warren,J, Warren,K, Wei,X, White,F, Williams,G, Willson,R, Wleczkyk,R, Wooden,H, Worley,K, Wright,D, Wright,K, Wu,J, Yakub,X, Yeon,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Zhao,S, Dunn,D, von Niedehausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished  
2 (bases 1 to 264606)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: KCMW  
Center clone name: CH230-66B6  
----- Summary Statistics  
Assembly program: Phrap, version 0.990329  
Consensus quality: 210747 bases at least Q40  
Consensus quality: 214941 bases at least Q30  
Consensus quality: 218032 bases at least Q20  
Estimated insert size: 219724; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOT: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 66491: contig of 66491 bp in length  
\* 66492 66591: gap of unknown length  
\* 66592 109095: contig of 42504 bp in length  
\* 109096 109195: gap of unknown length  
\* 109196 264606: contig of 15411 bp in length.

Location/Qualifiers  
1..264606  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-66B6"  
66592..67666  
/note="wgs contig"  
82489..84829  
/note="wgs contig"

BASE COUNT 56111 a 53481 c 54146 g 56734 t 44134 others

ORIGIN  
Query Match 28.2%; Score 35; DB 2; Length 264606;  
Best Local Similarity 100.0%; Pred. No. 4.9e-10;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GGATCCCTGCTTATGATGTCAAGCTGAAACAAAC 68  
|||||  
Db 158502 GGATCCCTGCTTATGATGTCAAGCTGAAACAAAC 158468

RESULT 40  
BC011127  
LOCUS BC011127 1125 bp mRNA linear ROD 16-APR-2003  
DEFINITION Mus musculus ring finger protein 7, mRNA (cDNA clone MGC:19076 IMAGE:4194107), complete cds.  
ACCESSION BC011127  
VERSION BC011127.1 GI:15029807  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schlier G.D., Altschul R.D., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhany S.J., Bosak S.A., McEwan P.J., Mckernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalske U., Smallos D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL  
MEDLINE  
22388257  
12477932  
2 (bases 1 to 1125)

REFERENCE  
TITLES  
Strausberg R.  
Direct Submission  
Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louesged, H., Kowals, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

FEATURES  
source  
1.1125  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:19076 IMAGE:4194107"  
/tissue\_type="Liver, normal. 5 month old male mouse."  
/clone\_1fb="NCI CGAP\_L19"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1.1125  
/gene="Rnf7"  
/note="synonym: SAG"  
/db\_xref="locusid:19823"  
/db\_xref="MGI:1337096"  
12.353  
/codon\_start=1  
/product="Rnf7 protein"

gene  
CDS

Query Match  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db  
11 GCCATCTGCAGGTCACAGTGAATGATGCTG 42  
162 GCCATCTGCAGGTCACAGTGAATGATGCTG 193

RESULT 41  
BD096966 1140 bp DNA linear PAT 27-AUG-2002  
DEFINITION  
SAG:apoptosis sensitivity gene.  
ACCESSION  
BD096966  
VERSION  
BD096966.1 GI:22642554  
KEYWORDS  
JP 2001526063-A/1.  
SOURCE  
unidentified  
unclassified  
ORGANISM  
unclassified.  
REFERENCE  
1 (bases 1 to 1140)  
Sun, Y.  
SAG:apoptosis sensitivity gene  
Patent: JP 2001526063-A 1 18-DEC-2001;  
WARNER LAMBERT CO  
OS Unidentified  
PN JP 2001526063-A/1  
PD 18-DEC-2001  
PF 15-DEC-1998 JP 2000525451  
PI 19-DEC-1997 US 60/068179, 11-SEP-1998 US 60/099840 PI  
YI SUN  
PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P17/02,A61P35/00,  
PC A61P39/06,  
PC A61P43/00,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC  
C12N5/10,C1201/68,  
PC G01N33/50,G01N33/68,C12N15/00,A61K37/02,C12N5/00 CC  
Strandedness: Double;  
CC Topology: Linear;  
CC /note = 'Mouse SAG';  
FH Key Location/Qualifiers  
FT CDS 17..355  
FT mat\_peptide 17..355  
FT misc\_feature 1..1140.  
Location/Qualifiers  
1.1140  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT  
302 a 224 c 287 g 327 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db  
11 GCCATCTGCAGGTCACAGTGAATGATGCTG 42  
167 GCCATCTGCAGGTCACAGTGAATGATGCTG 198

RESULT 42  
AF092877 1140 bp mRNA linear ROD 24-JUL-2001  
LOCUS  
AF092877  
DEFINITION  
Mus musculus zinc RING finger protein SAG mRNA, complete cds.  
ACCESSION  
AF092877

VERSION	KEYWORDS	AF092877.1	GI:4588031
SOURCE	ORGANISM	Mus musculus (house mouse)	
REFERENCE	AUTHORS	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 1140)	
TITLE		Duan,H., Wang,Y., Aviram,M., Swarcop,M., Loo,J.A., Bian,J., Tian,Y., Mueller,T., Bisgaier,C.L. and Sun,Y. SAG, a novel zinc RING finger protein that protects cells from apoptosis induced by redox agents	
JOURNAL	MEDLINE	Mol. Cell. Biol. 19 (4), 3145-3155 (1999)	
REFERENCE	AUTHORS	99182502 10082581 2 (bases 1 to 1140)	
TITLE		Sun,Y. Alterations of SAG mRNA in human cancer cell lines: requirement for the RING finger domain for apoptosis protection	
JOURNAL	MEDLINE	Carcinogenesis 20 (10), 1899-1903 (1999)	
REFERENCE	AUTHORS	99435944 10506102 3 (bases 1 to 1140)	
TITLE		Swarcop,M., Bian,J., Aviram,M., Duan,H., Bisgaier,C.L., Loo,J.A. and Sun,Y. Expression, purification, and biochemical characterization of SAG, a RING finger redox-sensitive protein	
JOURNAL	REFERENCE	Free Radical Biol. Med. 27, 193-202 (1999)	
REFERENCE	AUTHORS	4 (bases 1 to 1140) Swarcop,M., Wang,Y., Miller,P., Duan,H., Jatkoe,T., Madore,S.J. and Sun,Y. Yeast homolog of human SAG/ROC2/Rbx2/Hrt2 is essential for cell growth, but not for germination: chip profiling implicates its role in cell cycle regulation	
TITLE		Oncogene 19 (24), 2855-2866 (2000)	
JOURNAL	MEDLINE	20309684 10851089 5 (bases 1 to 1140)	
REFERENCE	AUTHORS	Duan,H., Tsvetkov,L.M., Liu,Y., Song,Y., Swarcop,M., Wen,R., Kung,H.F., Zhang,H. and Sun,Y. Promotion of S-phase entry and cell growth under serum starvation by SAG/ROC2/Rbx2/Hrt2, an E3 ubiquitin ligase component: association with inhibition of p27 accumulation	
TITLE		Mol. Carcinog. 30 (1), 37-46 (2001)	
JOURNAL	MEDLINE	21152847 11255262 6 (bases 1 to 1140)	
REFERENCE	AUTHORS	Sun,Y. Direct Submission Submitted (16-SEP-1998) Department of Molecular Biology, Parke-Davis, 2800 Plymouth Rd, Ann Arbor, MI 48105, USA	
FEATURES	source	location/Qualifiers 1..1140 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6xCBA" /db_xref="taxon:10090" /sex="female" /tissue_type="lung" /dev_stage="6-8 weeks" 17..358 /note="redox sensitive, metal binding; expression protects cells from apoptosis induced by redox compounds" /codon_start=1 /product="zinc RING finger protein SAG" /protein_id="AAD25961.1" /db_xref="GI:4588032" /translation="MADVEDDEPCEVLSHSGSAGSGDDKMFSLKKNAVAMSWDVECTCALICRVQVMDACIRCOAENKQEDCVVWGECHNSFHNCGMSLTVKONRRCPLCQODWVVRIGK"	
BASE COUNT		302 a 224 c 287 g 327 t	
Query Match		25.8%, Score 32, DB 10, Length 1140,	

[illegible]



VERSION	STV071691.1	GI:31187486
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 612)	
AUTHORS	Wade C.M., Kulbokas E.J. II, Kirby A.W., Zody M.C., Mullikin J.C., Lander E.S., Lindblad-Toh K. and Daly M.J.	
TITLE	The mosaic structure of variation in the laboratory mouse genome	
JOURNAL	Nature 420 (6915), 574-578 (2002)	
MEDLINE	22354684	
PUBMED	12466852	
COMMENT	Contact: Kerstin Lindblad-Toh Whitehead Institute for Biomedical Research, Center for Genome Research 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172521477 Fax: 6172580903 Email: Kers1@genome.wi.mit.edu Primer A: None Primer B: None STS size: 612 Protocol: WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 12951/Svimu, C3H/Heu, and BALB/cByJ. The WGS reads were placed uniquely on the MGS6v3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STS and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%. Location/Qualifiers 1..612 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57CH1/E1" /db_xref="taxon:10090" /map="10 22-541 63463278-63462744" /clone_id="C57CH11/E1" <1..>612 145 a 163 c 170 g 134 t	
FEATURES	source	
BASE COUNT	145 a	163 c 170 g 134 t
ORIGIN	STS	
Query Match	25.0%; Score 31; DB 11; Length 612;	
Best Local Similarity	100.0%; Pred. No. 1e-07;	
Matches	31; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
QY	16 CTGCAGGGTCAGGTGATGATGATGCTGCTT 46	
Db	379 CTGCAGGGTCAGGTGATGATGATGCTGCTT 409	
RESULT 45	AC103666 112321 bp DNA linear HTG 20-DEC-2002	
LOCUS	AC103666 Mus musculus clone RP23-397H13, LOW-PASS SEQUENCE SAMPLING.	
DEFINITION	AC103666	
ACCESSION	AC103666.2 GI:27311471	
VERSION	HTG; HTGS_PHASE0.	
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 112321)	
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Alt, A., Allen, N., Anderson, S., Barna, N., Baerlein, V., Boguski, L., Bouckgeater, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,	

Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,  
Lamazares, R., Landers, T., Lepocky, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 112321)

Birren, B., Nishpanu, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Borkgater, B.,  
Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Fato, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunhkhang, P., Pierre, N., Raymond, C., Retta, R.,  
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,  
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 20, 2002 this sequence version replaced gi:17149550.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L19615  
Center clone name: 397\_H\_13

NOTE: This record contains 89 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 1147: contig of 1147 bp in length  
\* 1148: gap of 100 bp  
\* 1248: contig of 1101 bp in length  
\* 2349: gap of 100 bp  
\* 2449: contig of 1063 bp in length

3512 3611: gap of 100 bp  
3612 4746: contig of 1135 bp in length  
4747 4846: gap of 100 bp  
4847 6030: contig of 1184 bp in length  
6031 6130: gap of 100 bp  
6131 7290: contig of 1160 bp in length  
7291 7390: gap of 100 bp  
7391 8577: contig of 1187 bp in length  
8578 8677: gap of 100 bp  
8679 9937: contig of 1260 bp in length  
9938 10037: gap of 100 bp  
10038 11192: contig of 1155 bp in length  
11193 11292: gap of 100 bp  
11293 12454: contig of 1162 bp in length  
12455 12554: gap of 100 bp  
12555 13718: contig of 1164 bp in length  
13719 13818: gap of 100 bp  
13819 15002: contig of 1184 bp in length  
15003 15102: gap of 100 bp  
15103 16234: contig of 1132 bp in length  
16235 16334: gap of 100 bp  
16335 17461: contig of 1127 bp in length  
17462 17561: gap of 100 bp  
17562 18650: contig of 1089 bp in length  
18651 18750: gap of 100 bp  
18751 19912: contig of 1162 bp in length  
19913 20012: gap of 100 bp  
20013 21199: contig of 1187 bp in length  
21200 21299: gap of 100 bp  
21300 22462: contig of 1163 bp in length  
22463 22562: gap of 100 bp  
22563 23720: contig of 1158 bp in length  
23721 23820: gap of 100 bp  
23821 25005: contig of 1185 bp in length  
25006 25105: gap of 100 bp  
25106 26265: contig of 1160 bp in length  
26266 26365: gap of 100 bp  
26366 27528: contig of 1163 bp in length  
27529 27628: gap of 100 bp  
27629 28788: contig of 1160 bp in length  
28789 28888: gap of 100 bp  
28889 30034: contig of 1146 bp in length  
30035 30134: gap of 100 bp  
30135 31236: contig of 1102 bp in length  
31237 31336: gap of 100 bp  
31337 32512: contig of 1176 bp in length  
32513 32612: gap of 100 bp  
32613 33813: contig of 1201 bp in length  
33814 33913: gap of 100 bp  
33915 35107: contig of 1194 bp in length  
35108 35207: gap of 100 bp  
35209 36431: contig of 1224 bp in length  
36432 36531: gap of 100 bp  
36532 37701: contig of 1170 bp in length  
37702 38967: contig of 1166 bp in length  
38968 39067: gap of 100 bp  
39069 40235: contig of 1168 bp in length  
40236 40335: gap of 100 bp  
40336 41492: contig of 1157 bp in length  
41493 41592: gap of 100 bp  
41593 42691: contig of 1099 bp in length  
42692 42791: gap of 100 bp  
42793 43970: contig of 1179 bp in length  
43971 44070: gap of 100 bp  
44071 45150: contig of 1080 bp in length  
45151 45250: gap of 100 bp  
45251 46347: contig of 1097 bp in length  
46348 46447: gap of 100 bp  
46449 47608: contig of 1161 bp in length  
47609 47708: gap of 100 bp  
47709 48881: contig of 1173 bp in length  
48882 48981: gap of 100 bp

48982 50240: contig of 1259 bp in length  
50241 50340: gap of 100 bp  
50341 51544: contig of 1204 bp in length  
51545 51644: gap of 100 bp  
51645 52828: contig of 1184 bp in length  
52829 52929: gap of 100 bp  
52929 54157: contig of 1229 bp in length  
54158 54257: gap of 100 bp  
54258 55450: contig of 1193 bp in length  
55451 55550: gap of 100 bp  
55551 56731: contig of 1181 bp in length  
56732 56831: gap of 100 bp  
56832 58019: contig of 1188 bp in length  
58020 58119: gap of 100 bp  
58120 59285: contig of 1166 bp in length  
59286 59385: gap of 100 bp  
59386 60525: contig of 1140 bp in length  
60526 60625: gap of 100 bp  
60626 61717: contig of 1092 bp in length  
61718 61817: gap of 100 bp  
61818 62991: contig of 1174 bp in length  
62992 63091: gap of 100 bp  
63092 64267: contig of 1176 bp in length  
64268 64367: gap of 100 bp  
64368 65552: contig of 1185 bp in length  
65553 65652: gap of 100 bp  
65653 66853: contig of 1201 bp in length  
66854 66953: gap of 100 bp  
66954 68145: contig of 1192 bp in length  
68146 68245: gap of 100 bp  
68246 69423: contig of 1178 bp in length  
69424 69523: gap of 100 bp  
69524 70737: contig of 1214 bp in length  
70738 70837: gap of 100 bp  
70838 71997: contig of 1160 bp in length  
71998 72097: gap of 100 bp  
72099 72554: contig of 1157 bp in length

Query Match  
Best Local Similarity 100.0%; Score 31; DB 2; Length 112321;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 CTCGAGGCTCCAGGTGATGATGCTGCTT 46  
Db 70515 CTCGAGGCTCCAGGTGATGATGCTGCTT 70545

RESULT 46  
AC103666/c  
LOCUS Mus musculus clone RP23-397H13, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC103666  
ACCESSION AC103666.2 GI:27311471  
VERSION HTG: HTGS PHASED.  
KEYWORDS Mus musculus (house mouse)  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 112321)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Darelid, K., Dewar, K., Diaz, U. S., Dodge, S., Fero, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K.,  
Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Liu, G.,

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Reta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Struss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travers, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 113321)

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L19615  
Center clone name: 397\_H\_13

Genome Center  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 20, 2002 this sequence version replaced gi:1719550.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information  
Center project name: L19615

Center clone name: 397\_H\_13

\* NOTE: This record contains 89 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 1147: contig of 1147 bp in length  
\* 1148 1247: gap of 100 bp  
\* 1248 2348: contig of 1101 bp in length  
\* 2349 2448: gap of 100 bp  
\* 2449 3511: contig of 1063 bp in length  
\* 3512 3611: gap of 100 bp  
\* 3612 4746: contig of 1135 bp in length  
\* 4747 4846: gap of 100 bp  
\* 4847 6030: contig of 1184 bp in length  
\* 6031 7290: contig of 1160 bp in length  
\* 7291 7390: gap of 100 bp

7391 8577: contig of 1187 bp in length  
\* 8578 8677: gap of 100 bp  
\* 8678 9937: contig of 1260 bp in length  
\* 9938 10037: gap of 100 bp  
\* 10038 11192: contig of 1155 bp in length  
\* 11193 11292: gap of 100 bp  
\* 11293 12454: contig of 1162 bp in length  
\* 12455 12554: gap of 100 bp  
\* 12555 13718: contig of 1154 bp in length  
\* 13719 13818: gap of 100 bp  
\* 13819 15002: contig of 1184 bp in length  
\* 15003 15102: gap of 100 bp  
\* 15103 16234: contig of 1132 bp in length  
\* 16235 16334: gap of 100 bp  
\* 16335 17461: contig of 1137 bp in length  
\* 17462 17561: gap of 100 bp  
\* 17562 18650: contig of 1089 bp in length  
\* 18651 18750: gap of 100 bp  
\* 18751 19912: contig of 1162 bp in length  
\* 19913 20012: gap of 100 bp  
\* 20013 21199: contig of 1187 bp in length  
\* 21200 21299: gap of 100 bp  
\* 21300 22462: contig of 1163 bp in length  
\* 22463 22562: gap of 100 bp  
\* 22563 23720: contig of 1158 bp in length  
\* 23721 23820: gap of 100 bp  
\* 23821 25005: contig of 1185 bp in length  
\* 25006 25105: gap of 100 bp  
\* 25106 26265: contig of 1160 bp in length  
\* 26266 27528: contig of 1163 bp in length  
\* 27529 27628: gap of 100 bp  
\* 27629 28788: contig of 1160 bp in length  
\* 28789 28888: gap of 100 bp  
\* 28889 30034: contig of 1146 bp in length  
\* 30035 30134: gap of 100 bp  
\* 30135 31236: contig of 1102 bp in length  
\* 31237 31336: gap of 100 bp  
\* 31337 32512: contig of 1176 bp in length  
\* 32513 32612: gap of 100 bp  
\* 32613 33813: contig of 1201 bp in length  
\* 33814 33913: gap of 100 bp  
\* 33914 35107: contig of 1194 bp in length  
\* 35108 35207: gap of 100 bp  
\* 35208 36431: contig of 1224 bp in length  
\* 36432 36531: gap of 100 bp  
\* 36532 37701: contig of 1170 bp in length  
\* 37702 37801: gap of 100 bp  
\* 37803 38967: contig of 1166 bp in length  
\* 38968 39067: gap of 100 bp  
\* 39068 40335: contig of 1168 bp in length  
\* 40336 40335: gap of 100 bp  
\* 40336 41492: contig of 1157 bp in length  
\* 41493 41592: gap of 100 bp  
\* 41593 42691: contig of 1099 bp in length  
\* 42692 42791: gap of 100 bp  
\* 42792 43970: contig of 1179 bp in length  
\* 43971 44070: gap of 100 bp  
\* 44071 45150: contig of 1080 bp in length  
\* 45151 45250: gap of 100 bp  
\* 45251 46347: contig of 1097 bp in length  
\* 46348 46447: gap of 100 bp  
\* 46448 47608: contig of 1161 bp in length  
\* 47609 47708: gap of 100 bp  
\* 47709 48881: contig of 1173 bp in length  
\* 48882 48981: gap of 100 bp  
\* 48982 50240: contig of 1259 bp in length  
\* 50241 50340: gap of 100 bp  
\* 50341 51544: contig of 1204 bp in length  
\* 51545 51644: gap of 100 bp  
\* 51645 52828: contig of 1184 bp in length  
\* 52829 52928: gap of 100 bp  
\* 52929 54157: contig of 1229 bp in length



```

misc_feature      3313. .6908      /note="assembly_fragment"
misc_feature      7009. 14414      /note="assembly_fragment"
misc_feature      14535. .32181     /note="assembly_fragment"
misc_feature      32282. .51167     /note="assembly_fragment"
misc_feature      51268. .72870     /note="assembly_fragment"
misc_feature      72971. .14992     /note="assembly_fragment"
misc_feature      150093. 187353     /note="assembly_fragment"
misc_feature      /note="assembly_fragment
vector_end:17
vector_slide:right"
BASE COUNT      60628 a 34855 c 36129 g 54836 t 905 others
ORIGIN

```

Query Match	25.0%	Score 31	DB 2	Length 187353
Best Local Similarity	100.0%	Pred. No. 1.5e+07		
Matches 31, Conservative	0	Mismatches 0	Indels 0	Gaps 0

Dy      11 GCATCTGCAGGGTCCAGGTGATCGATCCT 41

Dd      46893 GCATCTGCAGGGTCCAGGTGATCGATCCT 46922

RESULT	48
AC119145	
LOCUS	148290 bp
DEFINITION	Rattus norvegicus chromosome 1 clone RP2-4G14, WORKING DRAFT
	DNA linear HTG 11-JUN-2002

ACCESSION AC119145  
VERSION AC119145.1 GI:20303534  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULTPRP; HTGS\_DRAFT  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (pages 1 to 148290)  
 Taenzer, S., Monti, J., Gloeckner, G., Goesele, C., Baumgart, C.,  
 and ...

TITLE	JOURNAL
Rat chromosome 1 genomic sequence	Unpublished

REFERENCE  
AUTHORS  
2 (bases 1 to 148290)  
Taenzer, S. and Platzner, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (25-APR-2002) Genome Analysis, Institute of Molecular

REFERENCE  
AUTHORS  
Taenzer, S. and Platzter, M.  
3 (bases 1 to 148290)  
Biotechnolog. 11, Jena, Thüringia 07745, Germany

**TITLE** Direct Submission  
**JOURNAL** Submitted (11-JUN-2002) Genome Analysis, Institute of Molecular

COMMENT  
-----  
Biotechnology, Heutendstr. 11, Jena, Thuringia 07745, Germany  
-----  
Genome Center  
-----

Center: Institute of Molecular Biotechnology  
Center code: IMB

-----  
 web site: <http://genome.imb-jena.de/>  
 Contact: [gscj-submit@genome.imb-jena.de](mailto:gscj-submit@genome.imb-jena.de)  
 -----  
 Project Information

Center project name: rA08  
Center clone name: RP32-4G14

```
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
```

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329

Consensus quality: 140730 bases at least Q40  
Consensus quality: 143676 bases at least Q30

consensus quality: 145595 bases at least Q20  
Quality coverage: 8.49

.....

**Sequence Quality Assessment:** This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

```

** NOTE: This is a 'working draft' sequence. It currently
** consists of 10 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.

```

*	1	25366:	contig of 25364 bp in length
*	25365	25366:	gap of unknown length
*	25465	38093:	contig of 12629 bp in length
*	38094	38193:	gap of unknown length
*	38194	83197:	contig of 45604 bp in length
*	83198	83197:	gap of unknown length
*	83989	90673:	contig of 6776 bp in length
*	90674	90773:	gap of unknown length
*	90774	107588:	contig of 16815 bp in length
*	107589	107688:	gap of unknown length
*	107689	112170:	contig of 4482 bp in length
*	112171	112170:	gap of unknown length
*	112211	119805:	contig of 7535 bp in length
*	119806	119905:	gap of unknown length
*	119906	137949:	contig of 18044 bp in length
*	137950	138049:	gap of unknown length
*	138050	146503:	contig of 8454 bp in length
*	146504	146603:	gap of unknown length
*	146604	148290:	contig of 1687 bp in length

FEATURES	Location/Qualifiers
source	1. .148290

```
/organism="Rattus norvegicus"  
/mol_type="genomic DNA"
```

```
/db_xref="taxon:10116"  
/chromosome="1"
```

```

/clone="RP32-4G14"
/clone_lib="RP32 BAC Library"

```

	a	c	g	e	otherref
BASE COUNT	41779	31830	31870	41911	900
ORIGIN					

Query Match	24.2%	Score 30;	DB 2;	Length 148290;
-------------	-------	-----------	-------	----------------

Best Local Similarity 100.0%; Pred. No. 6.4e-07;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 TCGCCATCTGCAGGCTCCAGGTGATGAT 37

Db 17405 TGGCCATCTGCAGGTCACAGGTGATGGAT 17434

RESULT 49  
AC118520/c

LOCUS	AC118520	189541 bp	DNA	linear	HTG 20-NOV-2
DEFINITION	Rattus norvegicus clone CH230-34J21, *** SEQUENCING IN PROGRESS				

\*\*\*, 2 unordered pieces.  
AC118520  
AC118520

```

VERSION          AC118520.4  GI:25138134
KEYWORDS         HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED

```

SOURCE	ORGANISM
kattus norvegicus (Norway rat)	Rattus norvegicus
Poliovirus - Man	Polyomavirus

Eukaryota; metazoa; chordata; craniata; vertebrata; euteleostomi  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus

REFERENCE  
1 (bases 1 to 189541)  
Murray D, Mavris M, McElroy M, Lee R, Rhoades C, Adams C, Aldam T, Rattus.

Authors  
Muziny, D., Maite, M., Kere, A., Adams, C., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,



Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bismalo, K., Blair, J., Blankenburg, B., Blych, P., Brown, M., Bryant, K., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceesar, H., Center, A., Chacko, V., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Loulseged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Mijiga, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Wortley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

Unpublished  
Direct Submission  
2 (bases 1 to 198756)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (20-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 198756)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:24158510.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature

## table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBVZ
Center clone name: CH230-416B17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 187507 bases at least Q40
Consensus quality: 189483 bases at least Q30
Consensus quality: 190529 bases at least Q20
Estimated insert size: 192210; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 198756: contig of 198756 bp in length.
Location/Qualifiers
1..198756
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-416B17"
1..1088
/note="wgs end extension
clone_end:5p6"
misc_feature
1849..3549
/note="wgs end-extension
clone_end:5p6"
misc_feature
5730..6605
/note="clone boundary
clone_end:5p6
site:
end_sequence:B2197454"
complement(191645..192594)
/note="clone boundary
clone_end:T7
site_end:T7
end_sequence:B2197451"
197550..198756
/note="wgs end-extension
clone_end:T7"
BASE COUNT 51425 a 44460 c 43469 g 51998 t 7404 others
ORIGIN
Query Match 24.2%; Score 30; DB 2; Length 198756;
Best Local Similarity 100.0%; Pred. No. 6.5e-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 AGATGTCAGGCTGAAACAAACAAAGAGAC 76
DB 176453 AGATGTCAGGCTGAAACAAACAAAGAGAC 176482
Search completed: November 7, 2003, 13:29:07
Job time : 825.044 secs

```

**THIS PAGE BLANK (USPTO)**